

Measuring the impact of national guidelines: what methods can be used to uncover time-varying effects for healthcare evaluations?

Sarah Price^a, Xiaohui Zhang^b, Anne Spencer^a

^a*Medical School, University of Exeter, The United Kingdom*

^b*Business School, University of Exeter, The United Kingdom*

Abstract

We examine the suitability of three methods using patient-level data to evaluate the time-varying impacts of national healthcare guidelines. Such guidelines often codify progressive change and are implemented gradually; for example, National Institute for Health and Care Excellence (NICE) suspected-cancer referral guidelines. These were revised on 23/06/2015, to include more cancer symptoms and test results (“features”), partly reflecting changing practice. We explore the time-varying impact of guideline revision on time to colorectal cancer diagnosis, which is linked to improved outcomes in decision-analytic models. We included 11,842 patients diagnosed in 01/01/2006–31/12/2017 in the Clinical Practice Research Datalink with England cancer registry data linkage. Patients were classified by whether their first pre-diagnostic cancer feature was in the original guidelines (NICE-2005) or was added during the revision (NICE-2015-only). Outcome was diagnostic interval: time from first cancer feature to diagnosis. All analyses adjusted for age and sex. Two difference-in-differences analyses used either a Pre (01/08/2012–31/12/2014, n=2,243) and Post (01/08/2015–31/12/2017, n=1,017) design, or event-study cohorts (2006–2017 vs 2015) to estimate change in diagnostic interval attributable to official implementation of the revised guidelines. A semiparametric varying-coefficient model analysed the difference in diagnostic interval between the NICE groups over time. After model estimation, primary and broader treatment effects of guideline content and implementation were measured. The event-study difference-in-differences and the semiparametric varying-coefficient methods showed that shorter diagnostic intervals were attributable to official implementation of the revised guidelines. This impact was only detectable by pre-to-post difference-

in-differences when the pre/post periods were selected according to the estimation results from the varying-coefficient model. Formal tests of the parametric models, which are special cases of the semiparametric model, suggest that they are misspecified. We conclude that the semiparametric method is well suited to explore the time-varying impacts of guidelines codifying progressive change.

Research highlights

- Different methods can be used to measure the time-varying and treatment effects of guidelines
- We compare difference-in-differences and semiparametric varying-coefficient methods
- The 2015 UK national suspected-cancer referral guidelines are used as an example
- Two of the three methods identify time-varying effects
- There are no differences over time between men and women or for age

Keywords:

UK

Policy evaluation

General practitioners

Early identification of cancer

Time-varying impacts

Repeated cross-sectional data

Difference-in-differences

Semiparametric varying-coefficient method

1. Introduction

Many countries have bodies that develop national healthcare guidelines; for example, The National Institute for Health and Care Excellence (NICE) in England, The American College of Physicians in the USA, and the National Health and Medical Research Council in Australia. A guideline's impact is

likely to vary over time, dependent on structural and behavioural change, and may even start before publication if the guideline codifies improving clinical practice. As randomised controlled trials are often impracticable or unethical, there is considerable interest in methods applicable to observational data for quantifying the time-varying impacts of health-policy change (Craig et al., 2012). The current paper explores methods for capturing the time-varying effects of guidelines on health outcomes within repeated cross-sectional data with a control group. We illustrate these methods using the NICE suspected-cancer referral guidelines, first introduced in 2005 (NICE, 2005) and revised in 2015 (NICE, 2015). The guidelines aim to expedite cancer diagnosis for symptomatic patients attending primary care, because late diagnosis is associated with poor survival (Elliss-Brookes et al., 2012). The 2005 guidelines were based mainly on secondary care evidence, and included signs, symptoms and abnormal test results (collectively termed “clinical features”) suggesting a risk of undiagnosed cancer $\geq 5\%$ (NICE, 2005). The 2015 revision was based solely on primary-care evidence and added new clinical features. It also lowered the threshold risk to 3%, to liberalise investigation enough to expedite cancer diagnosis without overwhelming clinical services or increasing patient harms from over-investigation (NICE, 2015). The NICE guidelines apply in England, but are also endorsed in Wales and Northern Ireland, and are supported by The Independent Cancer Taskforce (Kumar, 2015).

One notable feature of these guidelines is that they codify the evidence base and best practice that develops over time, a characteristic not commonly encountered in intervention-evaluation studies. The most conservative measure of the effect of the guidelines is to measure the impact from the date of their official implementation, and this is the primary impact we measure. Broader effects can also be measured, such as those linked to guideline content and occurring before official implementation. This requires identifying a key moment in time from which we can explore change, which can be challenging.

We explore three methods that use a control group to assess the impact of revising the suspected-cancer guidelines on time to diagnosis of colorectal cancer – an outcome linked with improved survival in decision-analytic models (Tørring et al., 2013). The three methods are pre-to-post and event-study difference-in-differences, and a semiparametric varying-coefficient model.

The pre-to-post difference-in-differences method is widely applied in health-services research (Streeter et al., 2017), and was used to quantify the impact of the original 2005 NICE suspected-cancer referral guidelines on the timeliness of cancer diagnosis (Neal et al., 2014). The outcome measure was diagnostic interval, i.e. the time between first presentation of a cancer symptom to primary care and diagnosis (Weller et al., 2012). Diagnostic intervals were shorter in 2007–08 compared with 2001–02; however, this was only attributable to the new guidelines in two cancer sites (oesophagus and cervix) (Neal et al., 2014). It is possible that the guidelines had little impact, and/or that Neal et al.’s chosen control group did not capture the secular trend well. Alternatively, the pre-to-post difference-in-differences method may not be well equipped to measure time-varying changes in impact. This possibility is one focus of our paper.

The event-study difference-in-differences has been used to explore the phased impact of policy, by introducing a series of dummy variables into the difference-in-differences regression model, each indicating time relative to a new policy (Wing et al., 2018). These dummies are typically yearly or monthly, as used to examine the effects of US government school-desegregation policies on racial mix over time (Reber, 2005), and health-insurance mandates on mammography and breast cancer diagnosis (Bitler & Carpenter, 2016). This method can also be used to explore the time trends associated with pre-intervention anticipatory effects; for example, to examine the impact of changes in medical care insurance (Alpert, 2016; Brot-Goldberg et al., 2017) and hospital reform (Kolstad & Kowalski, 2012).

Finally, we consider the semiparametric varying-coefficient model (Li et al., 2002). Here, the pattern of time-varying effects is not restricted by any pre-specified parametric functional form, such as occurs with the common practice of adding interaction terms to a regression model. To our knowledge, the semiparametric varying-coefficient model is rarely applied in healthcare research. It has been used to explore time trends in the control and intervention arms within randomised controlled trials (Gilbert et al., 2019; Wasan et al., 2017), and the time-varying aspects of tumour characteristics on disease-free survival in breast cancer in an observational dataset (Natarajan et al., 2009).

There are three main ways in which the semiparametric model, applied to repeated cross-sectional data, provides additional insights over parametric

models in healthcare studies assessing the impact of interventions, particularly when the intervention is a guideline as we consider here. First, the method can account for the time-varying impact of all control variables (for example, age, sex, and intervention). Second, applying the semiparametric varying-coefficient model, it is possible to test the validity of the conventional parametric specifications, such as the pre-to-post and the event-study difference-in-differences methods. Third, once the varying-coefficient model has been run, treatment effects of guideline implementation, as well as the combining effects of both guideline content and subsequent implementation can be estimated. Given the novelty of the method, our aim is to illustrate the method’s potential and any assumptions required without being prescriptive about which of these features are ultimately used by researchers.

Our study adds to the existing literature in the following five ways. First, we describe the three methods and summarise their distinguishing characteristics. Second, we show how the semiparametric model can be used in preliminary analysis for data visualisation of the time-varying aspects of covariates. Third, we show that the pre-to-post and event-study difference-in-differences models are special cases of the semiparametric varying-coefficient model, enabling a formal test of the parametric models against the varying-coefficient model. Fourth, we show how the semiparametric method can be used in causal analysis to measure treatment effects. Finally, we report the first comparison of these difference-in-differences methods and the semiparametric varying-coefficient model to assess the impact of national guidelines in patient-level health records, using NICE referral guidelines for suspected colorectal cancer as an illustration.

2. Data

2.1. Study setting and inclusion criteria

The setting was primary care and the data sources were the Clinical Practice Research Datalink (CPRD), a dataset of observational, anonymised, patient records in UK primary care, with partial linkage to National Cancer Registration and Analysis Service (NCRAS) and Office for National Statistics (ONS) data (Herrett et al., 2015). Ethics approval by the Independent Scientific Advisory Committee (protocol 16_037) was obtained on 30/10/2017. The CPRD reduced temporarily in size from 2013 to 2017, so we anticipate fewer

Table 1: Features of colorectal cancer, by version of NICE suspected-cancer referral guidelines

In NICE 2005	Added in NICE 2015
Rectal bleeding	Abdominal pain
Iron-deficiency anaemia	Faecal occult blood
Change in bowel habit	Weight loss
Rectal mass	
Abdominal mass	

patients in the latter part of the study period, which was 1 January 2005 to 31 December 2017 inclusive. Patient inclusion criteria were: adults aged ≥ 18 years with an incident colorectal cancer diagnosis between 1 January 2006 and 31 December 2017, who were registered with the general practice for at least 1 year before their diagnosis. Patients were excluded if they: were diagnosed via screening; were registered with general practices in Scotland (where NICE suspected-cancer referral guidelines are not applicable); had multiple index cancers; had not attended their general practice in the year before diagnosis; or had not presented with any clinical features listed in NICE referral guidelines for suspected colorectal cancer in the year before diagnosis.

2.2. Variables

2.2.1. Diagnostic interval

We identified CPRD codes for features of undiagnosed colorectal cancer (i.e. abnormal test results and symptoms listed in NICE guidelines (Table 1)) in the year before diagnosis (NICE, 2005, 2015). Cancer diagnosis date was taken as the first diagnostic CPRD code, and is reliable (Boggon et al., 2013). Diagnosis date was validated by linked NCRAS data, where available, with the NCRAS data taking precedence if discrepant.

The outcome variable was diagnostic interval: the time between the first feature of cancer and diagnosis (Weller et al., 2012). Diagnostic interval was constrained to 1 year before diagnosis following standard practice, and was log-transformed as it has a right-skewed distribution (Hamilton et al., 2009).

2.2.2. Dummy variable for the NICE guidelines

For all three analysis methods, a dummy “ $NICE_{2015}$ ” variable was created by classifying all patients by their first recorded feature of cancer. The variable was set to 1 for patients with index feature(s) only listed in NICE-2015 guidelines and not previously in the 2005 guidelines, and 0 otherwise. In essence, the NICE-2015 group identifies patients captured by the expanded guidelines (Table 1).

2.2.3. Other control variables

The CPRD provided variables for sex, year of birth and region (England, Wales, Northern Ireland or Scotland). Age at diagnosis was estimated by assigning the birth date as 1 July of the year provided by the CPRD, and was entered into all analyses as a continuous variable. Patient-level Townsend Deprivation Index data (for England only), calculated using unadjusted 2001 census data, were provided by ONS.

In the main analysis, we control for patient sex and age at diagnosis. We have also conducted a series of sensitivity analyses by controlling for patient region (England versus Wales or Northern Ireland) and deprivation index.

3. Methods

3.1. Pre-to-post difference-in-differences

In the pre-to-post difference-in-differences method (Eq.(1)), the outcome of interest (diagnostic interval) is measured in four groups, *control* (NICE-2005) and *treated* (NICE-2015-only), *before* and *after* the treated group is managed according to the new guidelines. The effect attributable solely to official implementation of the new guidelines is estimated by the coefficient for the interaction term (δ_3) between the dummy variables d_T (indicating before or after revised guideline implementation) and $NICE_{2015}$ (indicating control or treated group) (Card & Krueger, 1994):

$$y = \delta_0 + \delta_1 NICE_{2015} + \delta_2 d_T + \delta_3 d_T \cdot NICE_{2015} + W\beta + u, \quad (1)$$

where vector y represents the outcomes of all observations, W is a matrix for other explanatory variables (for example, age and sex), and u is the standard residual error term with zero conditional mean.

In this study, we apply pre-to-post difference-in-differences to estimate the impact of implementing the revised NICE guidelines on log-transformed diagnostic interval. It is worth noting that the pre-to-post difference-in-differences model specified by Eq.(1) can only identify, under the so-called parallel assumption (see Appendix A for details), the average treatment effect of the new guideline’s implementation. It can neither accommodate the time-varying feature of this effect, nor capture the impact of the new guideline’s content before its official implementation.

We use data from patients diagnosed before (1 August 2012 – 31 December 2014) or after (1 August 2015 – 31 December 2017) revised guideline publication on 23 June 2015. Having comparable time periods before and after the guideline implementation date avoids seasonality effects. The analyses accounted for the correlation among individuals within the same general practices (clusters) and adjusted for age at diagnosis and sex. We checked for differences in the age and sex distributions between the Pre and Post periods. We use robust standard errors to account for heteroscedasticity and the models are run with $\log(\text{age})$ and with age-squared to check for non-linearity. In sensitivity analyses, we adjusted for deprivation and region (see Appendix C for detailed results).

3.2. The event-study difference-in-differences method

The event-study difference-in-differences (Eq.(2)) method estimates the time-varying effects of the revised guidelines:

$$y = \delta_0 + \delta_1 NICE_{2015} + \sum_{t=-T_l, t \neq 0}^{T_u} \gamma_t d_t + \sum_{t=-T_l, t \neq 0}^{T_u} \lambda_t d_t NICE_{2015} + W\beta + e \quad (2)$$

where d_t is a dummy variable representing the time period (which is annual here but can be any predetermined frequency) relative to the new guideline implementation, where $t = -T_l, -T_l + 1, \dots, 0, 1, \dots, T_u$. The period of guideline implementation is represented by $t=0$, with negative t values before implementation and positive t values after. $-T_l$ represents the earliest time period in the data, and T_u the most recent. $NICE_{2015}$ is defined as in Eq.(1), i.e. a dummy variable for treatment. For $t > 0$, λ_t measures the average treatment effect of the new guideline’s implementation in time period

t after the implementation. For $t < 0$, λ_t can reflect the trend of the anticipatory effect of the new guideline’s content before official implementation. e is the standard residual error term with zero conditional mean.

In this study we apply event-study difference-in-differences to estimate the time-varying impact of NICE guideline revision on log-transformed diagnostic interval. We use data from patients diagnosed with cancer between 1 January 2006 and 31 December 2017. We classified patients by the year of their index feature of cancer, using multiple dummy variables between 2005 ($T_l = 10$) and 2017 ($T_u = 2$) (see Eq.(2) above). Interaction terms between NICE category and each year dummy were sought, omitting year 2015 as the reference category. By applying this event-study difference-in-differences method, we identify under a variant parallel assumption (see Appendix A for details), the time-varying treatment effect of the guideline’s implementation on a yearly basis. It also visualises the pre-implementation anticipatory effect of the new guideline’s content (which are also yearly varying). For technical details, please see Appendix A. The analyses account for the correlation among individuals within the same general practices (clusters) and adjusts for age at diagnosis and sex. We check for differences in the age and sex distributions between the annual cohorts. In sensitivity analyses, we adjust for deprivation and region (see Appendix C).

3.3. The semiparametric varying-coefficient method

The semiparametric varying-coefficient method is represented by Eq.(3):

$$y = \delta_0(c_t) + \delta_1(c_t)NICE_{2015} + W\beta(c_t) + \epsilon, \quad (3)$$

where $NICE_{2015}$ and W are defined as in Eq. (1) and Eq. (2), and ϵ is the standard residual error term with zero conditional mean. In the varying-coefficient model (Eq.(3)), $NICE_{2015}$ and W are called regressors, and c_t is referred to as a covariate. c_t represents time relative to the new guideline’s implementation on 23 June 2015, with positive values for post-implementation time periods, negative values for pre-implementation time periods, and zero for time periods coinciding with guideline implementation. In particular, $c_t = -T_L, -T_L + 1, \dots, 0, 1, \dots, T_U$, where $-T_L$ and T_U represent the earliest and latest time points, respectively (equivalent to $-T_l$ and T_u in Eq.(2)). Note that, depending on the data frequency, c_t can be a discrete (month or year) or a continuous (day) variable. In this study, we take the unit of c_t

as a day, so we have $T_L=3825$ (1 January 2005) and $T_U=922$ (31 December 2017). In this model, we assume $E(u | NICE_{2015}, W, c_t) = 0$. The model can account for fixed effects associated with individual general practices (Feng et al., 2017; Rodriguez-Poo & Soberon, 2014), but this was not implemented here, owing to high level estimation demand and complexity.

We apply the semiparametric varying-coefficient method to estimate the impact of NICE guideline revision on the difference in mean log-transformed diagnostic interval between NICE-2015-only and NICE-2005 groups, for patients diagnosed with cancer between 1 January 2006 and 31 December 2017. Analyses were carried out using the ‘np’ package in R (Hayfield and Racine, 2008).

We illustrate three potential ways in which the semiparametric varying-coefficient method can be applied to obtain additional insights over the difference-in-differences parametric models.

First, we capture the time-varying (on a daily basis) difference in the diagnostic interval between the two groups (NICE-2005 and NICE-2015-only) over the entire time frame of the data. This allows visualisation of time-varying trends that may be associated with the new guidelines. The granularity of data analysis on a daily basis matches the typical reporting frequency of primary and secondary care patient records. Though the event-study difference-in-differences method also accounts for the time-varying impact of the new guidelines, the reporting frequency adopted is on a yearly basis, because considering shorter periods, such as per month or per day, would be cumbersome to implement using dummy variables and over-consume degrees of freedom (Wing et al., 2018). An additional feature of the semiparametric method is that it permits separate reporting of the time-varying coefficients of each regressor (for example, age, sex, and intervention). The corresponding results are presented in Section 4.3.1.

Second, we illustrate how the semiparametric varying-coefficient model can be used to test the functional forms imposed by the other two methods, i.e. the pre-to-post and the event-study difference-in-differences methods, to fit the data. In Section 3.4, we introduce the formal tests of the varying-coefficient model versus each of the two conventional difference-in-differences methods, respectively.

Finally, we illustrate how, once the semiparametric varying-coefficient model is run, the estimated coefficients can be used to measure the average treatment effect of guideline implementation itself on a particular day after the official implementation (i.e. $c_t > 0$). This is measured by the difference between the coefficient of $NICE_{2015}$ for the date of $c_t > 0$ and the coefficient for the date of guideline implementation (i.e. $\delta_1(c_t) - \delta_1(0)$). In addition, without re-running the model, we can measure broader treatment effects. In this example, it is interesting to explore the impact of the guideline's content, which we assume begins at time c_{t_0} . The impact is measured as $\delta_1(c_t) - \delta_1(c_{t_0})$, where $c_t > c_{t_0}$. This is challenging because there is great uncertainty around the actual date of c_{t_0} . To demonstrate the advantage of this method, we chose to measure the impact of the guideline's content from 30 June 2014, the date of a political announcement encouraging the recognition of early signs of cancer. Arguably, this announcement is weakly linked to the revised guideline's content. We can use this date to illustrate how the method captures the impact of the guideline's content, by subtracting the coefficient for $NICE_{2015}$ on the date of the announcement ($c_{t_0} = -357$) from the coefficient for $NICE_{2015}$ on any date *after* the political announcement ($c_t > -357$), i.e. $\delta_1(c_t) - \delta_1(-357)$. It should be noted that after guideline implementation (i.e. for $c_t > 0$), this measure captures the treatment effect of both the guideline's content and its implementation. The technical details on model estimation and average treatment effect identification are presented in Appendix B.

3.4. Model Comparison

In Eq.(3), the coefficients $\delta_0(c_t)$, $\delta_1(c_t)$, and $\beta(c_t)$ are all dependent on c_t . In contrast to the parametric model specifications in Eq.(1) and Eq.(2), the coefficients are not specified as a certain functional form; rather, unknown functional forms are assumed and estimated. In this sense, the parametric model specifications of Eq.(1) and Eq.(2) are two special cases of the semiparametric model, Eq.(3). In particular, Eq.(1) can be written as:

$$y = (\delta_0 + \delta_2 d_T) + (\delta_1 + \delta_3 d_T) NICE_{2015} + W\beta + u. \quad (4)$$

Therefore, when

$$\delta_0(c_t) = \begin{cases} \delta_0 & \text{if } c_t < 0 \\ \delta_0 + \delta_2 & \text{if } c_t \geq 0 \end{cases},$$

$$\delta_1(c_t) = \begin{cases} \delta_1 & \text{if } c_t < 0 \\ \delta_1 + \delta_3 & \text{if } c_t \geq 0 \end{cases},$$

and $\beta(c_t) \equiv \beta$, Eq.(3) collapses to Eq.(1). Similarly Eq.(2) can be written as:

$$y = \left(\delta_0 + \sum_{t=-T_l, t \neq 0}^{T_u} \gamma_t d_t \right) + \left(\delta_1 + \sum_{t=-T_l, t \neq 0}^{T_u} \lambda_t d_t \right) \cdot NICE_{2015} + W\beta + u. \quad (5)$$

In the same sense, it is a special case of Eq.(3), i.e.

$$\delta_0(c_t) = \begin{cases} \delta_0 & \text{if } c_t \text{ is in the time period of guideline implementation year} \\ \delta_0 + \gamma_t & \text{if } c_t \text{ is in the time period of } d_t = 1 \end{cases},$$

$$\delta_1(c_t) = \begin{cases} \delta_1 & \text{if } c_t \text{ is in the time period of guideline implementation year} \\ \delta_1 + \lambda_t & \text{if } c_t \text{ is in the time period of } d_t = 1 \end{cases},$$

and $\beta(c_t) \equiv \beta$.

This specification test can be used in isolation of any future features of the semiparametric model. In particular we note that that in this form the $\delta_1(c_t)$ from the semiparametric varying-coefficient model is contaminated by any pre-existing difference between treated and control groups (as indicated by δ_1 in Eq.(1) and Eq.(2)). While the estimand $\delta_1(c_t)$ is looking at differences between trends in the control (NICE-2005) and treated (NICE-2015-only) groups, rather than identifying a particular treatment effect, it is a useful first step in visualising the time-varying trends.

This comparison is drawn to show that the pre-to-post and event-study difference-in-differences model specifications are special cases of the semiparametric varying-coefficient model. This enables a formal test of the semiparametric varying-coefficient model versus the parametric models, using an approach proposed by Li and Racine (2010). The model specification test is outlined in more detail in Appendix B.3.

Importantly, this model specification test is distinct from any of the modifications needed to enable causal estimation using the semiparametric method, which are outlined in more detail in Appendix B.

4. Results

4.1. Study cohorts

The CPRD provided 25,011 patients with an incident colorectal cancer diagnosis between 1 January 2006 and 31 December 2017. Of these, 13,169 were excluded, mostly because they did not present with features of colorectal cancer ($n=9,693$) or because they resided in Scotland ($n=3,105$) where different guidelines apply. Patients diagnosed following screening ($n=52$), or with multiple index cancers ($n=52$), or who did not attend general practice ($n=4$) were also excluded, leaving 11,842 patients ($n=6,477$ men, 54.7%) in the analyses. Table 2 summarises the demographic characteristics of the patients included in the three different methods of analysis.

For the pre-to-post difference-in-differences analysis, there were fewer patients in the Post ($n=1,017$) than in the Pre ($n=2,243$) group. For the event-study difference-in-differences and semiparametric varying-coefficient models, the numbers of patients in the annual cohorts declined from 2013 to 2017 (Table 2). These falls are consistent with the known reducing size of the CPRD.

The percentages of NICE-2015-only and NICE-2005 patients was the same in the Pre and Post groups, and varied slightly year on year between 2005 and 2017 but no overall trend was apparent (Table 3).

There were no discernible differences in age distribution between the Pre and Post groups (see Figure 1), or between the annual cohorts (Table 2).

The sex distribution indicated a marginally greater proportion of men in the Post compared to the Pre group, and in years 2016–17 compared with 2005–15 (Table 2).

The assumption of parallel trends in the difference-in-differences models was explored by visual inspection of the data (see Figure C.1 in supplementary Appendix C). The general direction of change of mean diagnostic interval was broadly similar in the NICE-2015-only and NICE-2005 groups, apart from during the year of guideline revision (2015).

Table 2: Demographic characteristics of the patients included in pre-to-post and event-study difference-in-differences analyses, and in semiparametric varying-coefficient analysis (SPVC)

Analysis type	Period ^a	N (%)	Age ^b	% male
Pre-post diff-in-diff	Pre ^c	2,243 (68.8)	70.2 (13.1)	53.7
	Post ^d	1,017 (31.2)	68.8 (13.3)	58.4
	Total	3,260 (100)	69.8 (13.2)	55.2
Event-study diff-in-diff & SPVC				
	2005	299 (2.5)	70.3 (13.5)	53.5
	2006	1,239 (10.5)	70.1 (12.0)	53.0
	2007	1,213 (10.2)	70.5 (11.9)	53.8
	2008	1,174 (9.9)	70.7 (11.9)	54.4
	2009	1,191 (10.1)	70.3 (12.7)	54.2
	2010	1,213 (10.2)	70.7 (12.1)	57.4
	2011	1,148 (9.7)	69.5 (13.0)	53.8
	2012	1,131 (9.6)	70.0 (12.9)	55.3
	2013	964 (8.1)	70.3 (13.3)	54.9
	2014	837 (7.1)	70.4 (12.9)	51.3
	2015	653 (5.5)	70.3 (13.4)	54.1
	2016	486 (4.1)	68.6 (12.9)	61.3
	2017	294 (2.5)	68.2 (13.2)	59.2
	Total	11,842 (100)	70.1 (12.6)	54.7

Notes:

^a Period in which feature of possible cancer presented.

^b Age (in years) at diagnosis, mean (SD).

^c Pre: 1 Aug 2012 to 31 Dec 2014.

^d Post: 1 Aug 2015 to 31 Dec 2017.

Table 3: Numbers of patients by NICE group and study period in the pre-to-post and event-study difference-in-differences and in the semiparametric varying-coefficient models

Analysis type	Cohort	NICE-2005	NICE-2015-only	Total
Pre-post diff-in-diff	Pre ^a	1,331 (59.3)	912 (40.7)	2,243
	Post ^b	603 (59.3)	414 (40.7)	1,017
	Total	1,934	1,326	3,260
Event-study diff-in-diff & SPVC	2005	154 (51.5)	145 (48.5)	299
	2006	684 (55.2)	555 (44.8)	1,239
	2007	717 (59.1)	496 (40.9)	1,213
	2008	677 (57.7)	497 (42.3)	1,174
	2009	955 (55.0)	536 (45.0)	1,191
	2010	678 (55.9)	535 (44.1)	1,213
	2011	650 (56.6)	498 (43.4)	1,148
	2012	691 (61.1)	440 (38.9)	1,131
	2013	560 (58.1)	404 (41.9)	964
	2014	500 (59.7)	337 (40.3)	837
	2015	401 (61.4)	252 (38.6)	653
	2016	281 (57.8)	205 (42.2)	486
	2017	177 (60.2)	117 (39.8)	294
	Total	6,825 (57.6)	5,017 (42.4)	11,842

Notes:

^a Pre: 1 Aug 2012 to 31 Dec 2014.

^b Post: 1 Aug 2015 to 31 Dec 2017.

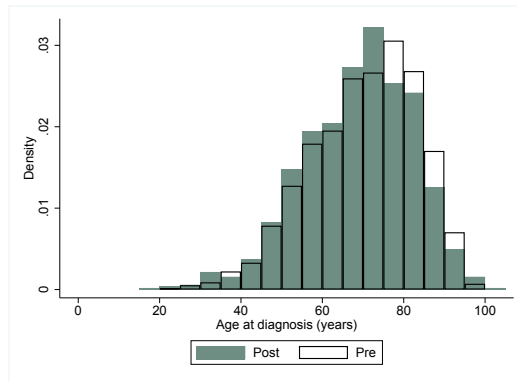


Figure 1: Distribution of age at diagnosis for patients in the main pre-to-post difference-in-differences analysis. Teal bars: Pre (n=2,243); black outlined bars: Post (n=1,017)

Table 4: Pre-to-post difference-in-differences regression analysis results

	Coefficient	95% CI	p
δ_0	3.789	3.569 to 4.009	<0.0001
δ_1	0.059	−0.046 to 0.164	0.271
δ_2	0.089	−0.009 to 0.187	0.075
δ_3	−0.139	−0.324 to 0.046	0.141
Age at diagnosis	0.001	−0.002 to 0.004	0.485
Sex (female vs male)	0.104	0.024 to 0.185	0.011

Notes:

See Eq.(1) for model specification:

δ_0 is mean log-transformed diagnostic interval for men at age 0 in the NICE-2005 group in the Pre cohort.

δ_1 is the difference in mean log-transformed diagnostic interval in the Pre cohort between the NICE-2015-only and NICE-2005 groups.

δ_2 is the Pre to Post difference in mean log-transformed diagnostic interval in the NICE-2005 group.

δ_3 is the Pre-to-Post change in mean log-transformed diagnostic interval attributable to NICE 2015.

Age at diagnosis was entered as a continuous variable.

4.2. Results from difference-in-differences methods

The pre-to-post difference-in-differences analysis results, adjusting for age and sex, are reported in Table 4. No secular trend in diagnostic interval (δ_2 : 0.089, 95% confidence interval −0.009 to 0.187, $p=0.075$), or impact of guideline revision (δ_3 : −0.139, −0.324 to 0.046, $p=0.141$) was found. Women have 10% longer diagnostic intervals than men (0.104, 0.024 to 0.185, $p=0.011$).

The event-study difference-in-differences analysis results are plotted in Figure 2 (see Table C.1 in supplementary Appendix C for model estimates). The trend of the anticipatory effect is to reduce diagnostic intervals prior to 2015. In 2017, there was moderate evidence that the impact of official guideline implementation in 2015 was to reduce diagnostic intervals by 30% (−0.307, 95% CI −0.616 to 0.003, $p=0.053$) (Figure 2).

Supplementary Appendix C reports sensitivity analyses, showing no non-linear effects of age (Table C.2), no heterogeneous effects of age between men and women (Tables C.3, C.4, Figure C.2), and no time-varying effects (Tables C.5, C.6).

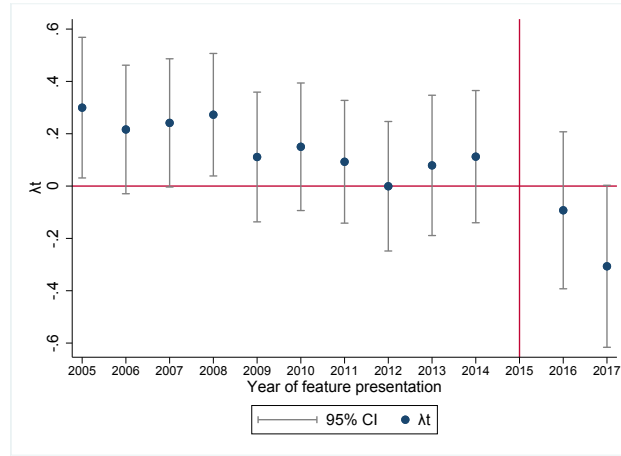


Figure 2: Data analysed using event-study difference-in-differences. Red vertical line denotes the year that NICE suspected-cancer guidelines were revised. Red horizontal line denotes no difference in diagnostic interval between NICE-2015-only and NICE-2005 groups. See Table C.2 in supplementary Appendix C for values

4.3. Results from varying-coefficient model

4.3.1. Data visualisation of time-varying effects

Figure 3 shows the percentage change in mean diagnostic interval (with 5% and 95% percentiles) between the NICE-2015-only and NICE-2005 groups over time estimated by the semiparametric varying-coefficient method. The percentage change oscillated with a period of approximately 1 year, and an amplitude of between 10 and 20 percentage points.

Between 1 July 2005 and 1 July 2008, NICE-2015-only patients had diagnostic intervals that were between 20% and 40% longer than those of NICE-2005 patients. From 1 July 2008 until shortly before the revised guidelines were introduced, the difference in diagnostic interval between the two groups reduced to zero. After the revised guidelines were introduced, NICE-2015-only patients had diagnostic intervals that were 20% to 30% shorter than those for NICE-2005 patients.

There are no suggestions of disparities in the data in terms of percentage change over time in mean diagnostic interval between men and women (Figure 4) or for age at diagnosis (Figure 5).

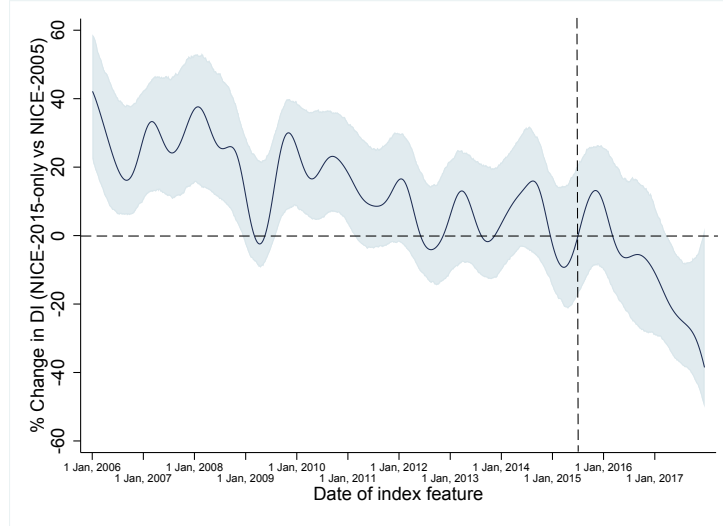


Figure 3: Mean diagnostic interval (DI) in NICE-2015-only patients versus NICE-2005 patients between 2006 and 2017. Data analysed using semiparametric varying-coefficient method. Black dashed vertical line denotes the introduction of revised suspected-cancer referral guidelines on 23 June 2015. Black dashed horizontal line denotes no difference in diagnostic interval between NICE-2015-only and NICE-2005 patients

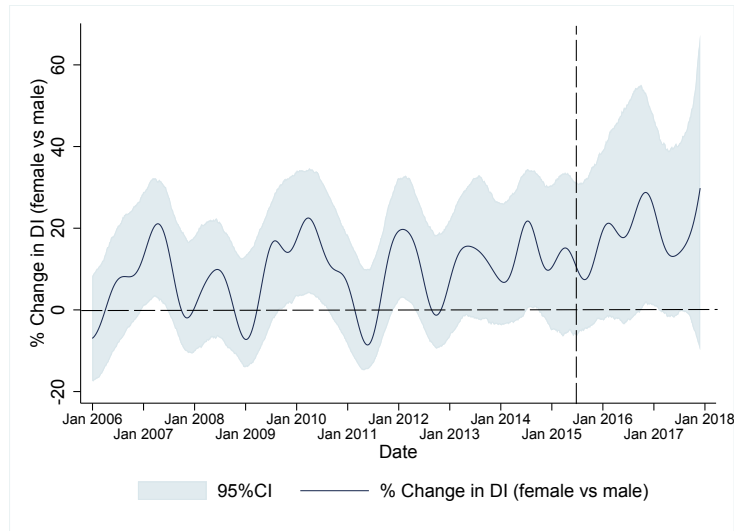


Figure 4: Trend in gender coefficient over time. Black dashed vertical line denotes the introduction of revised suspected-cancer referral guidelines on 23 June 2015. Black dashed horizontal line denotes no difference in diagnostic interval between women and men

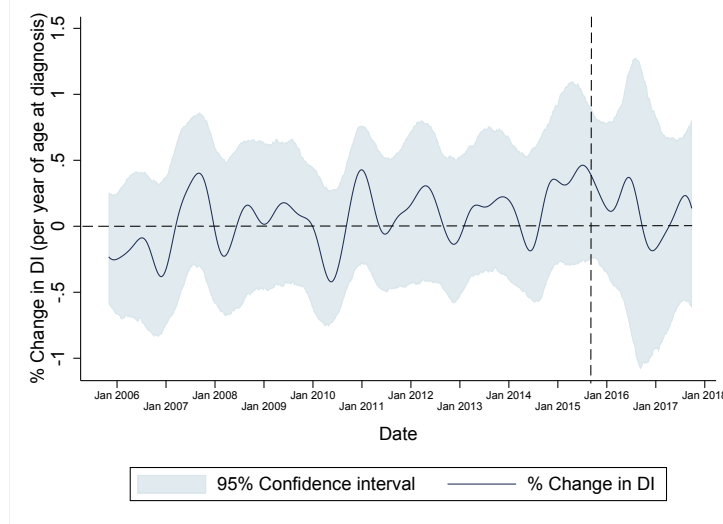


Figure 5: Trend in age coefficient over time. Black dashed vertical line denotes the introduction of revised suspected-cancer referral guidelines on 23 June 2015. Black dashed horizontal line denotes no difference in diagnostic interval with age

Figure 6 reports the trend in the intercept (mean log diagnostic interval) over time. With wide confidence intervals, there is no discernible change in absolute diagnostic interval over the whole study period.

By conducting the model specification test (see Appendix B.3), we can reject both the pre-to-post difference-in-differences model specification and the event-study difference-in-differences model specification, at the 1% significance level.

4.3.2. Treatment effect estimation

The primary, and most conservative, treatment effect to measure is that of *official guideline implementation*. The average treatment effect of official guideline implementation on a particular day after 23 June 2015 (i.e. $c_t > 0$) is estimated by the difference between the coefficients for $NICE_{2015}$ on the date of interest and on the date of guideline implementation (i.e. $\delta_1(c_t) - \delta_1(0)$).

We also illustrate how the broader effects of guidelines can be measured, by selecting the date of 30 June 2014 ($c_t \geq -357$), when the then Secretary of State for Health, Jeremy Hunt, encouraged GPs to improve their recognition

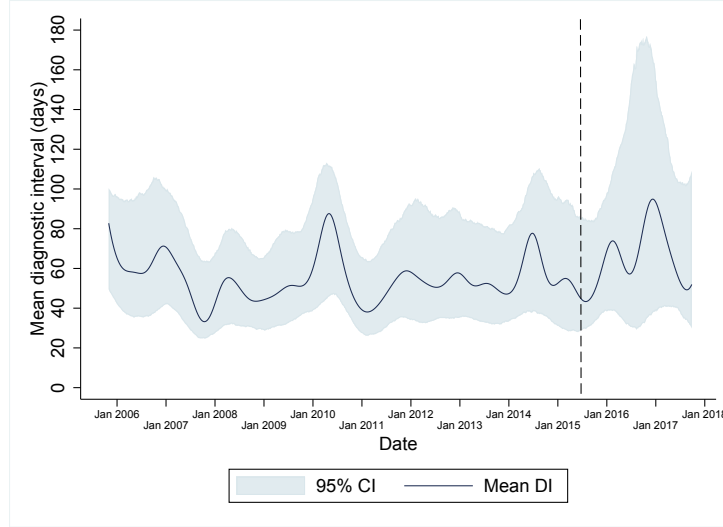


Figure 6: Trend in the intercept. Black dashed vertical line denotes the introduction of revised suspected-cancer referral guidelines on 23 June 2015

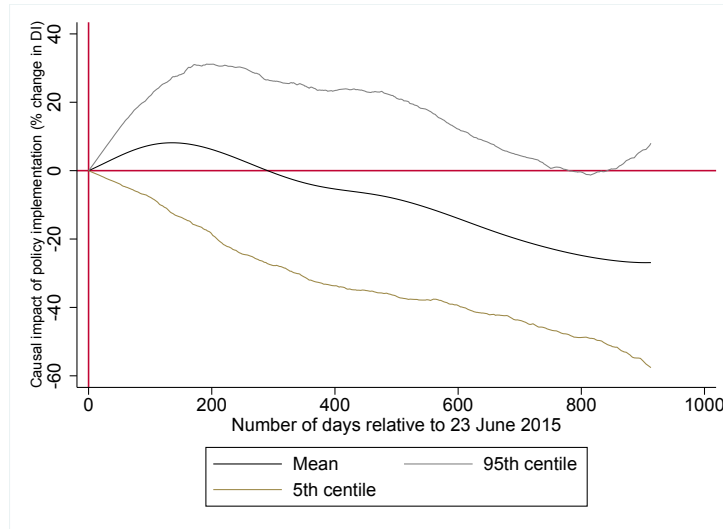


Figure 7: Treatment effect of guideline implementation on 23 June 2015

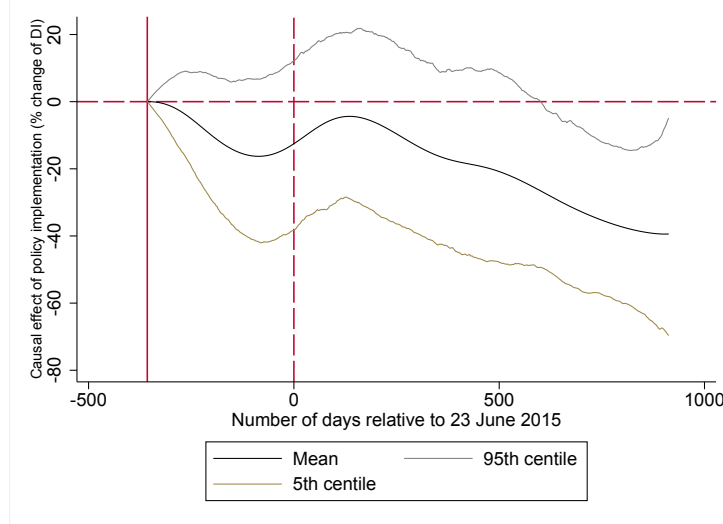


Figure 8: Treatment effect of guideline content and implementation, assuming guideline content impact began on 30 June 2014 (solid vertical red line)

of the early signs of cancer. This type of treatment effect can be measured by $\delta_1(c_t) - \delta_1(-375)$.

The treatment effect of *official policy implementation* is plotted in Figure 7. The results indicate a trend towards a positive treatment effect of the guidelines to reduce diagnostic intervals, but the 90% confidence interval is wide and the upper limit remains greater than zero, suggesting considerable uncertainty. It is unfortunate that the CPRD was reducing in size over this time, which may contribute to this uncertainty.

The broader treatment effects are plotted in Figure 8, which shows an immediate impact of guideline content to reduce diagnostic interval, followed by a rebound rise and then subsequent and sustained fall after official guideline implementation on 23 June 2015.

4.4. Robustness checks on pre-to-post difference-in-differences analysis

In a series of robustness checks, we examined the effect of changing the Pre and Post periods. The results are presented in supplementary Appendix C. Changing the start time in four 10-day increments within the confines of

the data availability did not alter the model estimates (Table C.7). Varying the start time and the width (Pre: 1/7/2013–31/7/2013; Post: 1/7/2017–31/7/2017) to examine immediate impacts had no effect. However, analyses including 1-year Pre (2013) and Post (2017) windows sandwiching guideline implementation in 2015 reported a significant reduction in diagnostic interval associated with guideline implementation (δ_3 -0.393 , 95% confidence interval -0.693 to -0.093 , $p=0.010$) (Table C.8).

Next, we selected Pre (16/3/2014–31/12/2014) and Post (16/3/2017–31/12/2017) periods informed by the results of the semiparametric varying-coefficient analyses. The coefficient comparing diagnostic interval between NICE-2015-only and NICE-2005 groups became significantly and consistently negative on 16 March 2017. The results showed a significant impact of NICE guideline revision on diagnostic interval (δ_3 -0.476 , 95% CI -0.807 to -0.146 , $p=0.005$) (Table C.8).

Finally, to enable comparison of all three models, we included all observations in the original data set, and set the cut-off time as the date of guideline revision (23 June 2015). This showed a significant impact of NICE guideline revision (δ_3 -0.215 , 95%CI -0.374 to -0.055 , $p=0.008$) (Table C.8).

4.5. *Sensitivity analyses for unobserved confounding*

In sensitivity analyses (see supplementary Appendix C), we examined the effects of potential confounders other than age and sex. Region is a potential confounder because of geographical variation in the recognition and reporting of cancer symptoms (Maja Niksic et al., 2016) and in the dissemination of guidelines in England versus Wales or Northern Ireland. In sensitivity analyses for all three models, a binary variable for region (England versus Wales or Northern Ireland) was included as a regressor. The main results were unaffected, but diagnostic intervals are shorter in England than in Wales or Northern Ireland (pre-to-post difference-in-differences: -0.218 , 95%CI -0.306 to -0.129 , $p<0.0001$, Table C.9; event-study: -0.309 , -0.359 to -0.259 , $p<0.0001$, Table C.10). The semiparametric varying-coefficient model reported a trend for this discrepancy to reduce over time (Figure C.3).

Deprivation may act as a confounder through effects on the recognition and reporting of cancer symptoms and consulting behaviour (M. Niksic et al., 2015). In sensitivity analyses, all three methods analysed England-only

data (owing to Townsend data availability), including a binary explanatory variable for deprivation (most deprived quintiles 4 and 5, versus least deprived quintiles 1–3). The coefficient for deprivation was not significant in the pre-to-post difference-in-differences (Table C.11), but in the event-study difference-in-differences (Table C.12) and semiparametric varying-coefficient (Figure C.4) analyses, there was evidence that diagnostic interval tended to be longer with increasing deprivation.

5. Discussion

This is the first study, to our knowledge, that compares methods to measure the time-varying impacts of national guidelines using repeated, cross-sectional patient-level data.

Our data source was the CPRD – one of the largest database of anonymised patient records in the world, and a widely used resource for primary care research and epidemiology studies (Herrett et al., 2015). Robust methods were used for case identification in the electronic medical record (Watson et al., 2017). Our sample sizes, of several hundred patients per year, were adequate for measuring the secular trend and divergences from it. Limitations relating to the temporary reduction in CPRD size towards the end of the study are discussed below.

The event-study difference-in-differences and the semiparametric varying-coefficient methods showed that revising the suspected-cancer referral guideline was associated with shorter diagnostic intervals for colorectal cancer. In contrast, these impacts were only detectable by the pre-to-post difference-in-differences method when the start dates and widths of the pre and post periods were carefully selected according to the varying-coefficient model estimates (Section 4.4). Furthermore, the semiparametric tests of coefficient functional specification suggest that the pre-to-post and event-study difference-in-differences models are misspecified.

This study builds on econometric literature (Li et al., 2002; Li and Racine, 2010) developed to explore time-varying effects. Given that the semiparametric method has not yet been widely applied in health services research, our aim has been to illustrate the method’s potential without being prescriptive on its use. The strengths of the semiparametric varying-coefficient

method are now summarised briefly. The time-varying impacts are not restricted by any specified parametric functional form, such as the common practice of adding interaction terms to regression models. The semiparametric model is able to uncover more granular changes in time trends compared with the difference-in-differences methods. The model can be used to explore the time-varying effects of confounders such as age and sex, and is useful for the visualisation of data generally. The method can also be used to test the specification of difference-in-differences models, and can be adapted for use in causal analysis.

Finally, we note that a potential challenge for all methods is how to capture any behavioural changes arising before guideline implementation – owing to the difficulty of identifying when these impacts begin. The conventional difference-in-differences methods typically approach this problem by only measuring the impact from the date of implementation. To estimate impact before that requires re-running these models, assuming different reference periods. In contrast, the semiparametric method needs to be run only once – after which it is possible to identify key dates from which we can explore the broader treatment effects of guidelines.

We conclude that the semiparametric varying-coefficient method offers useful features that provide additional insights into the time-varying aspects of repeated cross-sectional data.

5.1. Limitations

The reducing size of the CPRD in 2013–17 occurred as contributing general practices moved away from the CPRD’s required IT system, Vision. The similar demographics of the annual cohorts in the current study suggests that any variation in the numbers of CPRD practices over time did not introduce bias. We carried out more detailed examination for associated bias in a separate study (Price et al., under review by *Cancer Epidemiology*), but found none.

For causal analysis, all three models are based on different versions of the parallel trends assumption as outlined in Appendix A and Appendix B. Meeting the parallel trends assumption is essential for reliable estimates because it ensures that the only difference between the groups is exposure to the

guidelines (Abadie, 2005). All the methods require careful selection of the control group to ensure that it accurately represents secular trends, which are then assumed to hold for the treated group. It could be argued that the semiparametric method makes the strongest parallel trends assumption, being on a daily basis. For this reason, we have not been prescriptive in how researchers should apply the semiparametric varying-coefficient model, and show that its other features are valuable in their own right.

The challenges of meeting the parallel trends assumption have led to literature looking at, among other things, synthetic controls (Kreif et al., 2016; O’Neill et al., 2016) which weight the combination of groups used as controls against which the treatment group is compared. Propensity score weighting within difference-in-differences can be used to weight the composition of treatment pre, treatment post, control pre, and control post groups to ensure a balanced set of characteristics which largely overcomes the misspecification due to observed covariates (Stuart et al., 2014).

5.2. Comparison with emerging literature

The methods that we have reviewed apply a variant of the regression difference-in-differences design to take account of time-varying effects. Callaway and Sant’Anna (2019) have recently proposed an alternative to the regression difference-in-differences method that applies a generalised propensity score method, using probit or logit, to sample a proportional effect for the treated group. This method was developed to accommodate multiple treatment effects arising from interventions that: (1) start at different time points and (2) can influence multiple time periods. A useful addition of their paper is a proposed procedure to aggregate multiple treatment effects – which is particularly helpful to measure the effect of complex staggered implementation. But this feature is not required here as the NICE suspected-cancer guidelines were launched nationwide within a short space of time. In principle, the Callaway and Sant’Anna (2019) method can be used to capture the time-varying treatment effect, but, similar to the event-study difference-in-differences method, the time periods are restricted to a relatively small number and would be very difficult to implement with a high-frequency time variable, such as daily data as in our application.

5.3. Future research

In the current paper we have explored the abilities of different models to capture the time-varying effects of new guidelines after their implementation. It is also important to measure the pre-implementation effects of new guidelines that codify gradual improvements in clinical practice over time. To do this requires identifying a key moment in time, for example the release date of underpinning research findings, from which we can measure change. While we have illustrated how the semiparametric method is able to do this in principle, a full examination of this issue was beyond the scope of the current paper, and will be the subject of further research.

In the current paper we are able to explore the impact of guidelines on diagnostic interval. To inform decision analytic models of early cancer diagnosis there is also need to explore other outcomes, such as primary care interval (i.e. time from first cancer feature to referral), referrals and stage at diagnosis. These outcomes are all essential parts of the evidence base needed to link expedited cancer diagnosis to improved cancer outcomes. Such research will require the CPRD dataset to be linked with Hospital Episode Statistics data to follow the patient from primary to secondary care, which is beyond the scope of this current study.

We have also shown how the semiparametric model can explore the impact the trends in other variables, such as age and sex. One extension of this feature would be to apply the semiparametric model to explore whether the impact of the new guidelines is heterogeneous across age at diagnosis and gender. This extension would enable researchers to explore the potential for unintended consequences of the guidelines on the distribution of health across observed confounders – which would be important had the time-varying trends in age and sex alone depicted change.

References

- Abadie, A. (2005). Semiparametric difference-in-differences estimators. *Rev Economic Stud*, 72, 1-19.
- Alpert, A. (2016). The anticipatory effects of Medicare Part D on drug utilization. *J Health Econ*, 49, 28-45.

Bitler, M.P., & Carpenter, C.S. (2016). Health insurance mandates, mammography, and breast cancer diagnoses. *Am Econ J Econ Policy*, 8, 39-68.

Boggon, R., van Staa, T.P., Chapman, M., Gallagher, A.M., Hammad, T.A., & Richards, M.A. (2013). Cancer recording and mortality in the General Practice Research Database and linked cancer registries. *Pharmacoepidemiol Drug Safety*, 22, 168-175.

Brot-Goldberg, Z.C., Chandra, A., Handel, B.R., & Kolstad, J.T. (2017). What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. *Q J Econ*, 123, 1261-1318.

Callaway, B., Sant’Anna, P.H.C., Difference-in-differences with multiple time periods, working paper March 2019.

Card, D., & Krueger, A.B. (1994). Minimum-wages and employment – a case-study of the fast-food industry in New-Jersey and Pennsylvania. *Am Econ Rev*, 84, 772-793.

Craig, P., Cooper, C., Gunnell, D., Haw, S., Lawson, K., Macintyre, S., et al. (2012). Using natural experiments to evaluate population health interventions: new MRC guidance. *J Epidemiol Commun Health*, 66, 1182-1186.

Elliss-Brookes, L., McPhail, S., Ives, A., Greenslade, M., Shelton, J., & Hiom, S. (2012). Routes to diagnosis for cancer – determining the patient journey using multiple routine data sets. *Br J Cancer*, 107.

Feng, G., Gao, J., Peng, B., & Zhang, X. (2017). A varying-coefficient panel data model with fixed effects: Theory and an application to US commercial banks. *J Econom*, 196, 68-82.

Gilbert, R., Brown, M., Rainford, N., Donohue, C., Fraser, C., Sinha, A., et al. (2019). Antimicrobial-impregnated central venous catheters for prevention of neonatal bloodstream infection (PREVAIL): an open-label, parallel-group, pragmatic, randomised controlled trial. *Lancet Child Adolesc Health*, 3, 381-390.

- Hamilton, W., Lancashire, R., Sharp, D., Peters, T.J., Cheng, K., & Marshall, T. (2009). The risk of colorectal cancer with symptoms at different ages and between the sexes: a case-control study. *BMC Med*, 7, 17.
- Hayfield, T., & Racine, J.S. (2008). Nonparametric Econometrics: The np Package. *J Stat Softw*, 27(5), 1-32.
- Herrett, E., Gallagher, A.M., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., et al. (2015). Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*, 44, 827-836.
- Kolstad, J.T., & Kowalski, A.E. (2012). The Impact of Health Care Reform on Hospital and Preventive Care: Evidence from Massachusetts. *J Public Econ*, 96, 909-929.
- Kreif, N., Grieve, R., Hangartner, D., Turner, A.J., Nikolova, S., & Sutton, M. (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Econ*, 25, 1514-1528.
- Kumar, H. (2015). Achieving world-class cancer outcomes: a strategy for England 2015-2020. London: The Independent Cancer Taskforce.
- Li, Q., Huang, C.J., Li, D., & Fu, T.T. (2002). Semiparametric smooth coefficient models. *J Bus Econ Stat*, 20, 412-422.
- Li, Q., & Racine, J. S. (2010). Smooth varying-coefficient estimation and inference for qualitative and quantitative data. *Econometric Theory*, 26(6), 1607-1637.
- Natarajan, L., Pu, M.Y., Parker, B.A., Thomson, C.A., Caan, B.J., Flatt, S.W., et al. (2009). Time-varying effects of prognostic factors associated with disease-free survival in breast cancer. *Am J Epidemiol*, 169, 1463-1470.
- Neal, R.D., Din, N.U., Hamilton, W., Ukoumunne, O.C., Carter, B., Stapley, S., et al. (2014). Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. *Br J Cancer*, 110, 584-592.

NICE (2005). Referral Guidelines for Suspected Cancer CG27. London: National Institute for Health and Clinical Excellence.

NICE (2015). Guidelines for Suspected cancer NG12. London: National Institute for Health and Care Excellence.

Niksic, M., Rachet, B., Duffy, S.W., Quaresma, M., Møller, H., & Forbes, L.J.L. (2016). Is cancer survival associated with cancer symptom awareness and barriers to seeking medical help in England? An ecological study. *Br J Cancer*, 115, 876-886.

Niksic, M., Rachet, B., Warburton, F.G., Wardle, J., Ramirez, A.J., & Forbes, L.J.L. (2015). Cancer symptom awareness and barriers to symptomatic presentation in England – are we clear on cancer? *Br J Cancer*, 113, 533-542.

O’Neill, S., Kreif, N., Grieve, R., Sutton, M., & Sekhon, J.S. (2016). Estimating causal effects: considering three alternatives to difference-in-differences estimation. *Health Serv Outcomes Res Methodol*, 16, 1-21.

Price, S.J., Spencer, A., Zhang, X., Ball, S., Lyratzopoulos, G., Mujica-Mota, R., et al. (2020) Trends in time to cancer diagnosis around the period of changing national guidance on referral of symptomatic patients: A serial cross-sectional study using UK electronic healthcare records. *Cancer Epidemiol* (in review)

Reber, S.J. (2005). Court-ordered desegregation – Successes and failures integrating American schools since Brown versus Board of Education. *J Hum Resour*, 40, 559-590.

Rodriguez-Poo, J.M., & Soberon, A. (2014). Direct semi-parametric estimation of fixed effects panel data varying coefficient models. *Econom J*, 17, 107-138.

Streeter, A.J., Lin, N.X., Crathorne, L., Haasova, M., Hyde, C., Melzer, D., et al. (2017). Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. *J Clin Epidemiol*, 87, 23-34.

Stuart, E.A., Huskamp, H.A., Duckworth, K., Simmons, J., Song, Z., Chernew, M., et al. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health Serv Outcomes Res Methodol*, 14, 166-182.

Torring, M.L., Frydenberg, M., Hansen, R.P., Olesen, F., & Vedsted, P. (2013). Evidence of increasing mortality with longer diagnostic intervals for five common cancers: a cohort study in primary care. *Eur J Cancer*, 49, 2187-2198.

Wasan, H.S., Gibbs, P., Sharma, N.K., Taieb, J., Heinemann, V., Riche, J., et al. (2017). First-line selective internal radiotherapy plus chemotherapy versus chemotherapy alone in patients with liver metastases from colorectal cancer (FOXFIRE, SIFLOX, and FOXFIRE-Global): a combined analysis of three multicentre, randomised, phase 3 trials. *Lancet Oncol*, 18, 1159-1171.

Weller, D., Vedsted, P., Rubin, G., Walter, F.M., Emery, J., Scott, S., et al. (2012). The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer*, 106, 1262-1267.

Wing, C., Simon, K., & Bello-Gomez, R.A. (2018). Designing difference in difference studies: best practices for public health policy research. *Annu Rev Public Health*, 39, 453-469.

Appendix

This document includes three sections. In the first two sections, we discuss in detail the treatment effect identification and measurement by different methods, i.e. the pre-to-post difference-in-differences (DID) method, the event-study DID method and the semiparametric varying-coefficient method, the estimation method for the varying-efficient model and the formal test for the varying-coefficient model versus the parametric model specifications. The description and discussion is under the particular context of our application, i.e. to examine the impact of new national guidelines in 2015 on the diagnostic interval for colorectal cancer using repeated cross-sectional data from CPRD between 1 January 2006 and 31 December 2017. In the last section (Appendix C), we present the extra results from a series of sensitivity analyses and robustness checks.

Appendix A. Treatment Effect Identification from the Pre-to-post DID Method and the Event-study DID Method

In this section, we discuss the treatment effect identification from the pre-to-post DID method and the event-study DID method. First, we consider the pre-to-post DID method, i.e.

$$y = \delta_0 + \delta_1 NICE_{2015} + \delta_2 d_T + \delta_3 d_T \cdot NICE_{2015} + W\beta + u, \quad (\text{A.1})$$

where d_T is a dummy variable indicating before ($d_T = 0$) or after ($d_T = 1$) the new guideline implementation. By using this pre-to-post DID method, we aim to identify the average treatment effect of the new guideline implementation for the treated sub-population (ATT).

Let $y^1(i, d_T)$ represent the potential diagnostic interval of individual i , when the first recorded feature of cancer observed during time period of d_T is one of the NICE-2015-only features, and $y^0(i, d_T)$ indicates the potential diagnostic interval of individual i , when his/her first recorded feature of cancer observed during time period of d_T is one of the NICE-2005 features. The ATT of the new guideline implementation conditional on W can be defined as $ATT = E[y^1(i, 1) - y^0(i, 1) | W_i, NICE_{2015,i} = 1]$.

To reduce notation, we drop the argument i , the indicator of individuals. To identify the ATT, we have to rely on the so-called parallel assumption, as stated below:

Assumption A.1. $E[y^0(1) - y^0(0)|W, NICE_{2015} = 1] = E[y^0(1) - y^0(0)|W, NICE_{2015} = 0]$.

Assumption A.1 states that the average difference in the diagnostic intervals before and after the new guideline implementation for the control and treated groups, conditional on the other observed factors W , would have been the same in the absence of the new guideline implementation. Under Assumption A.1, if the pre-to-post DID method of Eq(A.1) is applied, the ATT can be identified and measured by δ_3 in Eq (A.1), because

$$\begin{aligned}
ATT &= E[y^1(1) - y^0(1)|W, NICE_{2015} = 1] \\
&= E[y^1(1) - y^0(0) + y^0(0) - y^0(1)|W, NICE_{2015} = 1] \\
&= E[y^1(1) - y^0(0)|W, NICE_{2015} = 1] - E[y^0(1) - y^0(0)|W, NICE_{2015} = 1] \\
&= E[y^1(1) - y^0(0)|W, NICE_{2015} = 1] - E[y^0(1) - y^0(0)|W, NICE_{2015} = 0] \\
&= \left[(\delta_0 + \delta_1 + \delta_2 + \delta_3 + W\beta) - (\delta_0 + \delta_1 + W\beta) \right] - \left[(\delta_0 + \delta_2 + W\beta) - (\delta_0 + W\beta) \right] \\
&= \delta_3
\end{aligned} \tag{A.2}$$

Next, we consider the event-study DID method, as follows

$$y = \delta_0 + \delta_1 NICE_{2015} + \sum_{t=-T_L, t \neq 0}^{T_U} \gamma_t d_t + \sum_{t=-T_L, t \neq 0}^{T_U} \lambda_t d_t NICE_{2015} + W\beta + e, \tag{A.3}$$

where d_t is a dummy variable indicating the year relative to the new guideline implementation year (i.e. 2015), with $t > 0$ for years after 2015, $t = 0$ for 2015, and $t < 0$ for a year before 2015. By applying the event-study DID method, we are able to identify the average treatment effect, conditional on the observed factors W , of the new guideline implementation in a certain year for the treated sub-population ($ATT(t|W)$).

Let $y^1(i, t)$ represent the potential diagnostic interval of individual i , when the first recorded feature of cancer observed during the year t is one of

the NICE-2015-only features, and $y^0(i, t)$ indicates the potential diagnostic interval of individual i , when his/her first recorded feature of cancer observed during the year t is one of the NICE-2005 features. The $ATT(t|W)$ of the new guideline implementation conditional on W can be defined as $ATT(t|W) = E[y^1(i, t) - y^0(i, t)|W_i, NICE_{2015,i} = 1]$.

Following the same practice, from now on we drop the argument i . In order to identify $ATT(t|W)$, we have to assume a variant version of the parallel assumption. In particular, it is as follows:

Assumption A.2. $E[y^0(t) - y^0(0)|W, NICE_{2015} = 1] = E[y^0(t) - y^0(0)|W, NICE_{2015} = 0]$.

It is assumed that the average difference in the diagnostic intervals between the new guideline implementation year and any year after the implementation for the control and treated groups, conditional on the other observed factors W , would have been the same in the absence of the new guideline implementation. Under Assumption A.2, by applying the event-study DID method of Eq(A.3), the $ATT(t|W)$ can be identified and measured by λ_t ($t > 0$) in Eq (A.3). It can be shown that

$$\begin{aligned}
ATT(t|W) &= E[y^1(t) - y^0(t)|W, NICE_{2015} = 1] \\
&= E[y^1(t) - y^0(0) + y^0(0) - y^0(t)|W, NICE_{2015} = 1] \\
&= E[y^1(t) - y^0(0)|W, NICE_{2015} = 1] - E[y^0(t) - y^0(0)|W, NICE_{2015} = 1] \\
&= E[y^1(t) - y^0(0)|W, NICE_{2015} = 1] - E[y^0(t) - y^0(0)|W, NICE_{2015} = 0] \\
&= \left[(\delta_0 + \delta_1 + \gamma_t + \lambda_t + W\beta) - (\delta_0 + \delta_1 + W\beta) \right] - \left[(\delta_0 + \gamma_t + W\beta) - (\delta_0 + W\beta) \right] \\
&= \lambda_t
\end{aligned} \tag{A.4}$$

Appendix B. Technical details for the semiparametric varying-coefficient model

This appendix contains: (a) how to identify and measure the average treatment effect on the treated at a particular time point by the semiparametric varying-coefficient model, and (b) how to estimate the semiparametric varying-coefficient model using the kernel estimation method.

Appendix B.1. The treatment effect identification and calculation

Consider the semiparametric varying-coefficient model specified in this paper:

$$y = \delta_0(c_t) + \delta_1(c_t)NICE_{2015} + W\beta(c_t) + \epsilon, \quad (B.1)$$

As stated in the main text, c_t represents the day relative to the new guideline implementation, i.e. 23 June 2015. In particular, $c_t > 0$ indicates a post-official-implementation day, $c_t = 0$ is the official implementation day, and $c_t < 0$ represents a pre-official-implementation day. Under a variant parallel trend assumption, the estimated model coefficients can be used to identify the average treatment effect of the revised guideline at c_t (conditional on W and related to a particular event) on the diagnostic interval of patients whose first recorded feature of cancer was one of the NICE-2015-only features, i.e. the average treatment effect for the treated group at time point c_t . This is indicated as $ATT(c_t|W)$.

As mentioned in the paper, we can identify different types of treatment effect of the revised guideline, including the treatment effects of official guideline implementation and of the guideline's content. To measure these effects, we need to identify a specific date on which we believe they began. For the official guideline implementation, we assume this to be 23 June 2015, i.e. $c_t = 0$ defined in our model. For guideline content, we illustrate the method by selecting the date of a political announcement related to early signs of cancer made on 30 June 2014, i.e. $c_t = -357$, by the Secretary of State for Health.

We can identify the treatment effect of the revised guidelines starting from specific dates of interest, c_0 . For dates $c_t > c_0$, let $y^1(i, c_t)$ represent the potential diagnostic interval of individual i , when his/her first recorded feature of cancer observed at time c_t is one of the NICE-2015-only features, and $y^0(i, c_t)$ indicates the potential diagnostic interval of individual i , when his/her first recorded feature of cancer observed at time c_t is one of the NICE-2005 features. The treatment effect of new guideline implementation on the diagnostic interval for individual i at time c_t ($c_t > c_0$) that is relevant to the particular occurring at c_0 can be naturally defined as $y^1(i, c_t) - y^0(i, c_t)$. The identification problem for this treatment effect is that, for any particular individual i , we do not observed both $y^1(i, c_t)$ and $y^0(i, c_t)$. Owing to the impossibility of computing the individual treatment effect, researchers often

measure the average treatment effect for the treated, conditional on some observed characteristics. In this study, we define the average treatment effect of the guideline related to a particular event at c_t ($c_t > c_0$) conditional on W as $ATT_{c_o}(c_t|W) = E[y^1(i, c_t) - y^0(i, c_t)|W_i, NICE_{2015,i} = 1]$.

To reduce notation, from now on we drop the individual argument i . In order to identify $ATT_{c_o}(c_t|W)$, we need to assume:

Assumption B.1. $E[y^0(c_t) - y^0(c_o)|W, NICE_{2015} = 1] = E[y^0(c_t) - y^0(c_o)|W, NICE_{2015} = 0]$.

Assumption B.1 is a variant of the conventional parallel assumption and it is crucial for identifying the treatment effect of guidelines related to a particular event using this semiparametric varying-coefficient model. It states that, conditional on the regressors W , the average outcome (i.e. diagnostic interval) for the treated group ($NICE_{2015} = 1$) would have followed a path parallel to that of the control group ($NICE_{2015} = 0$) from the time that the particular event of interest ($c_t = c_0$) occurred until time c_t ($c_t > c_0$), in the absence of the treatment (i.e. no event related to the guideline revision happened).

Under Assumption B.1, by applying the semiparametric varying-coefficient model, the conditional average treatment effect $ATT_{c_o}(c_t|W)$ can be expressed and calculated as:

$$\begin{aligned}
ATT_{c_0}(c_t|W) &= E[y^1(c_t) - y^0(c_t)|W, NICE_{2015} = 1] \\
&= E[y^1(c_t) - y^0(c_0) + y^0(c_0) - y^0(c_t)|W, NICE_{2015} = 1] \\
&= E[y^1(c_t) - y^0(c_0)|W, NICE_{2015} = 1] - \\
&\quad E[y^0(c_t) - y^0(c_0)|W, NICE_{2015} = 1] \\
&= E[y^1(c_t) - y^0(c_0)|W, NICE_{2015} = 1] - \\
&\quad E[y^0(c_t) - y^0(c_0)|W, NICE_{2015} = 0] \\
&= \left[(\delta_0(c_t) + \delta_1(c_t) + W\beta(c_t)) - (\delta_0(c_0) + \delta_1(c_0) + W\beta(c_0)) \right] - \\
&\quad \left[(\delta_0(c_t) + W\beta(c_t)) - (\delta_0(c_0) + W\beta(c_0)) \right] \\
&= \delta_1(c_t) - \delta_1(c_0)
\end{aligned} \tag{B.2}$$

Appendix B.2. Kernel Estimation

In general, Eq.(B.1) can be written more compactly as

$$y_i = X_i'\theta(Z_i) + u_i, \quad i = 1, \dots, n, \tag{B.3}$$

where X_i is a column vector indicating all regressors of the i th individual, with the first element of X_i is 1, i.e. the intercept term; Z_i is the covariate (in our model, Eq.(3), it is the time of the i th individual's first observed cancer feature relative to the revised guideline implementation in 2015); $\theta(Z_i)$ is the coefficient of regressors X_i , which is allowed to vary across Z_i ; y_i and u_i are the i th individual's diagnostic interval and error term, respectively; and n is the sample size. In this study, we apply the kernel estimation method (Li and Racine, 2010) to estimate Eq.(B.3). In particular, as Z_i is continuous in our study, we use the kernel given by

$$W_h(Z_i, Z_j) = \frac{1}{h} \omega \left(\frac{Z_j - Z_i}{h} \right), \tag{B.4}$$

where $\omega(\bullet)$ is a symmetric univariate density function; for example, we adopt a univariate standard normal density function; and $0 < h < \infty$ is the smoothing parameter for Z . Let S denote the whole set of possible values for Z . For any $z \in S$, from Eq.(B.3) it can be shown that

$$\theta(z) = [E(X_i X_i' | Z_i = z)]^{-1} [E(X_i Y_i | Z_i = z)] \tag{B.5}$$

Therefore, by the local constant kernel method, $\theta(z)$ is estimated as

$$\hat{\theta}(z) = \left[n^{-1} \sum_{i=1}^n X_i X_i' W_h(Z_i, z) \right]^{-1} \left[n^{-1} \sum_{i=1}^n X_i y_i W_h(Z_i, z) \right] \quad (\text{B.6})$$

When h is sufficiently large, then $\omega((Z_i - z)/h) \rightarrow \omega(0)$, which is constant and unrelated to Z . Thus Z is completely smoothed out and $\hat{\theta}(z)$ does not depend on Z . In this sense, the magnitude of h indicates the relevance of Z , such that, if Z is relevant, as $n \rightarrow \infty$, the smoothing parameter h will converge to 0 (Li and Racine, 2010).

We use a cross-validation method to choose the optimal smoothing parameter h . In particular, the optimal h minimises

$$CV(h) = \sum_{i=1}^n \left(y_i - X_i' \hat{\theta}_{-i}(Z_i) \right)^2, \quad (\text{B.7})$$

where

$$\hat{\theta}_{-i}(Z_i) = \left[n^{-1} \sum_{\substack{j=1, \\ j \neq i}}^n X_j X_j' W_h(Z_j, Z_i) \right]^{-1} \left[n^{-1} \sum_{\substack{j=1, \\ j \neq i}}^n X_j y_j W_h(Z_j, Z_i) \right]$$

is the leave-one-out kernel estimator of $\theta(Z_i)$.

Appendix B.3. Testing for coefficient functional specification

The pre-to-post and event-study DID models are special cases of the semi-parametric varying-coefficient model. This enables a formal test of the parametric models by testing the parametric specification of $\theta(Z)$ using an approach proposed by Li and Racine (2010).

Specifically, let $\theta_0(Z)$ denote the coefficient vector in Eq.(4) or Eq.(5), which is a parametric function of Z . We test if $\theta(Z)$ is of a parametric function form as defined by $\theta_0(Z)$ over the whole set of possible values for Z . If it is not, we should avoid using the parametric specification of Eq.(4) or Eq.(5), i.e. a model specified by Eq.(1) or Eq.(2), due to its misspecification.

To conduct this model specification test, we formulate the null hypothesis as $H_0 : \Pr(\theta(Z) = \theta_0(Z)) = 1$, i.e. the probability of varying-coefficient model's nonparametrically specified coefficient $\theta(Z)$ being equal to the parametrically specified coefficient $\theta_0(Z)$ of the pre-to-post or the event-study difference-in-differences model is 1. Following Li and Racine (2010), the test statistic is given by

$$\hat{I}_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1, j \neq i}^n X_i' X_j \hat{u}_i \hat{u}_j W_{\hat{h}}(X_i, X_j), \quad (\text{B.8})$$

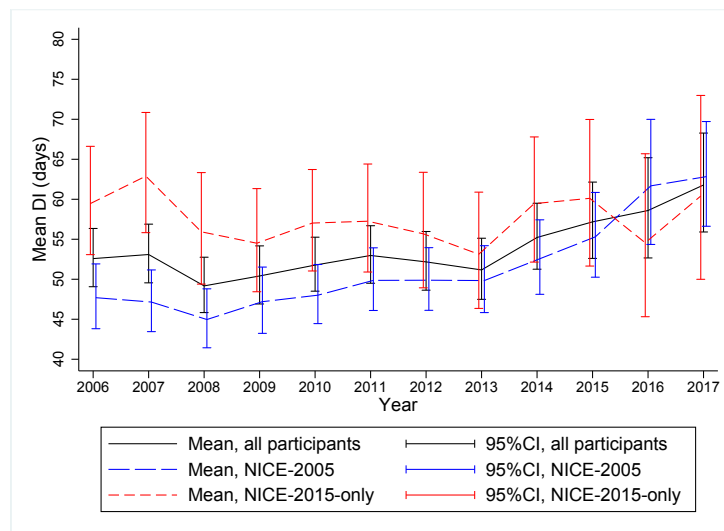
where \hat{h} is the optimal smoothing parameter chosen by the cross-validation method for the varying-coefficient model, and \hat{u}_i is the estimated error from the parametric model, i.e. $\hat{u}_i = y_i - X_i' \hat{\theta}_0(Z_i)$. Under H_0 , we have

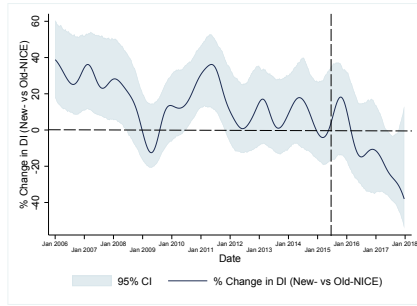
$$\hat{T}_n \stackrel{\text{def}}{=} n \sqrt{\hat{h}} \hat{I}_n / \hat{\sigma}_0 \xrightarrow{d} N(0, 1) \quad (\text{B.9})$$

where $\hat{\sigma}_0^2 = 2n^{-2} \sum_{i=1}^n \sum_{j \neq i}^n (X_i' X_j \hat{u}_i \hat{u}_j W_{\hat{h}}(X_i, X_j))^2$. A nonparametric bootstrap procedure to approximate the null distribution of \hat{T}_n proposed by Li and Racine (2010) is applied in this study.

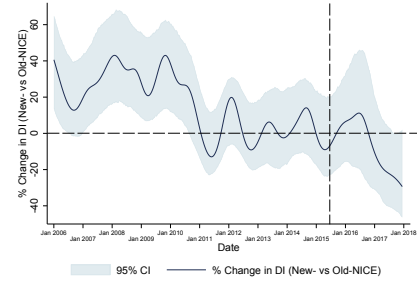
Appendix C. Robustness Check and Sensitivity Analysis

Figure C.1: Mean diagnostic interval (95% confidence interval) by year, for NICE-2015-only (red dashed line) and NICE-2005 (blue long-dashed line) patients, and for both groups combined (black solid line)

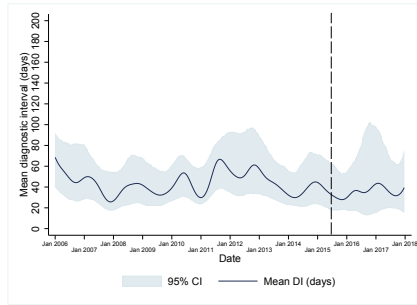




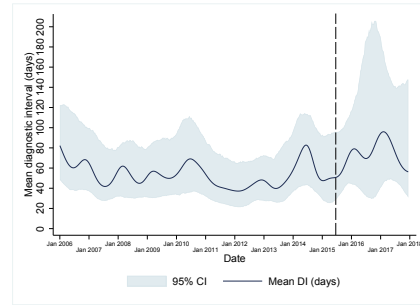
(a) New vs Old NICE: M



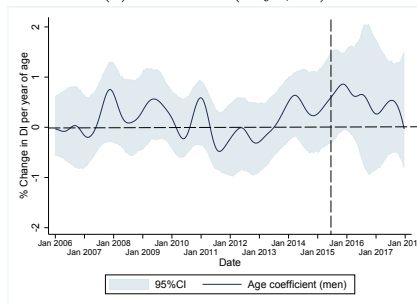
(b) New vs Old NICE: F



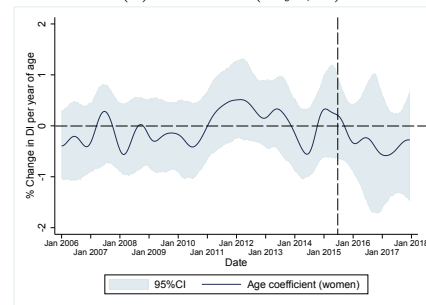
(c) Mean DI (days, M)



(d) Mean DI (days, F)

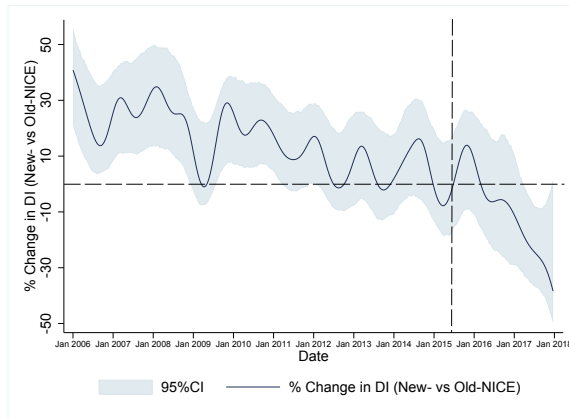


(e) Age coefficient: M

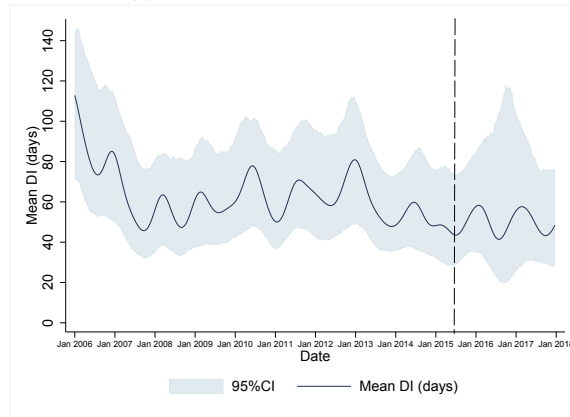


(f) Age coefficient: F

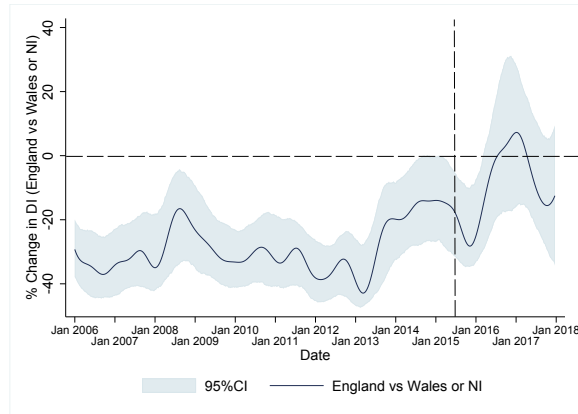
Figure C.2: Semiparametric varying-coefficient analyses of diagnostic interval from 1 Jan 2006 to 31 December 2017, for men (M) and women (F) separately



(a) NICE-2015 vs NICE-2005

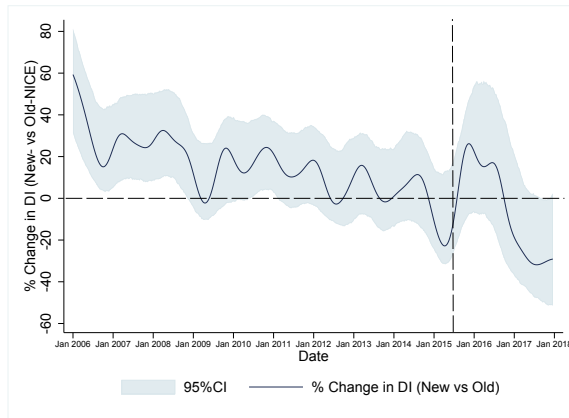


(b) Mean DI (days)

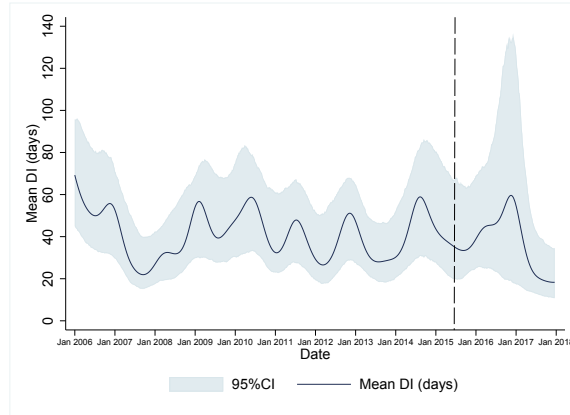


(c) England vs Wales or NI

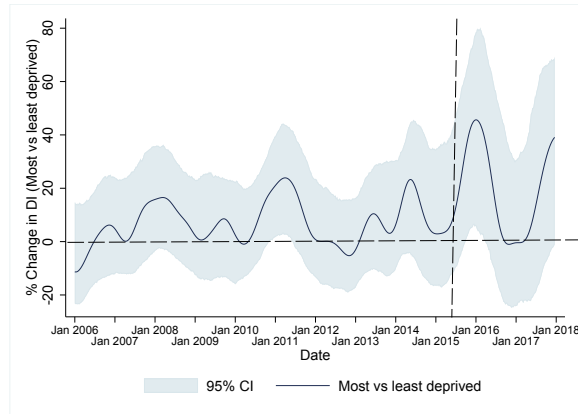
Figure C.3: Semiparametric varying-coefficient analyses of diagnostic interval from 1 Jan 2006 to 31 December 2017, adjusting for region (England vs Wales or Northern Ireland)



(a) NICE-2015 vs NICE-2005



(b) Mean DI (days)



(c) Most vs least deprived

Figure C.4: Semiparametric varying-coefficient analyses of diagnostic interval in England from 1 Jan 2006 to 31 December 2017, adjusting for deprivation quintile (most deprived quintiles 4 and 5 versus least deprived quintiles 1 to 3)

Table C.1: Results of event-study difference-in-differences analysis (see Figure 2)^{a b}

t (years relative to 2015)	Coefficient	Value	95% CI	P
	δ_0	3.88	3.727 to 4.042	<0.0001
	δ_1	0.007	-0.190 to 0.204	0.943
-10	γ_{-10}	0.831	0.674 to 0.988	<0.0001
	λ_{-10}	0.300	0.031 to 0.568	0.029
-9	γ_{-9}	-0.076	-0.215 to 0.062	0.277
	λ_{-9}	0.216	-0.029 to 0.462	0.084
-8	γ_{-8}	-0.138	-0.267 to -0.009	0.035
	λ_{-8}	0.241	-0.004 to 0.486	0.053
-7	γ_{-7}	-0.161	-0.288 to -0.035	0.013
	λ_{-7}	0.273	0.039 to 0.507	0.022
-6	γ_{-6}	-0.121	-0.255 to 0.012	0.074
	λ_{-6}	0.111	-0.137 to 0.359	0.379
-5	γ_{-5}	-0.067	-0.198 to 0.064	0.317
	λ_{-5}	0.150	-0.094 to 0.394	0.227
-4	γ_{-4}	-0.066	-0.190 to 0.058	0.297
	λ_{-4}	0.093	-0.142 to 0.327	0.437
-3	γ_{-3}	-0.017	-0.146 to 0.112	0.795
	λ_{-3}	-0.001	-0.248 to 0.247	0.996
-2	γ_{-2}	-0.125	-0.261 to 0.011	0.072
	λ_{-2}	0.079	-0.189 to 0.347	0.563
-1	γ_{-1}	0.001	-0.134 to 0.136	0.987
	λ_{-1}	0.112	-0.140 to 0.365	0.383
+1	γ_{+1}	0.091	-0.058 to 0.241	0.232
	λ_{+1}	-0.092	-0.393 to 0.207	0.545
+2	γ_{+2}	-0.008	-0.157 to 0.140	0.911
	λ_{+2}	-0.306	-0.616 to 0.003	0.052
	Age	0.001	-0.001 to 0.002	0.525
	Sex (f vs m)	0.090	0.046 to 0.134	<0.0001

^a Post-estimation link tests did not reveal any problems with the data specification of the difference-in-differences model.

^b No evidence of unobserved confounding was found using the omitted variable test ($p=0.9744$).
 δ_0 is mean log-transformed diagnostic interval for men aged 0 in the NICE-2005 group in 2015.
 δ_1 is the difference in mean log-transformed diagnostic interval between the NICE-2015-only and NICE-2005 groups in 2015.

γ_t is the difference in mean log-transformed diagnostic interval in NICE-2005 groups between year t and 2015.

λ_t is the difference in mean log-transformed diagnostic interval in year t compared with 2015 that is attributable to NICE 2015.

Table C.2: Robustness check for non-linear effect of age in the difference-in-differences regression

Estimand	Including age-squared			Including log(age)		
	Coeff.	95% CI	p	Coeff.	95% CI	p
δ_0	4.168	3.349 to 4.998	<0.0001	3.636	2.901 to 4.372	<0.0001
δ_1	0.055	-0.050 to 0.160	0.305	0.056	-0.048 to 0.161	0.292
δ_2	0.162	0.064 to 0.260	0.001	0.162	0.064 to 0.260	0.001
δ_3	-0.071	-0.253 to 0.112	0.449	-0.071	-0.254 to 0.111	0.443
Age	-0.010	-0.035 to 0.015	0.421	n/a	n/a	n/a
Age-squared	0.00009	-0.0001 to 0.0003	0.365	n/a	n/a	n/a
Log(age)	n/a	n/a	n/a	0.058	-0.115 to 0.232	0.506
Sex (f vs m)	0.123	0.046 to 0.200	0.02	0.127	0.050 to 0.204	0.001

δ_0 is mean log-transformed diagnostic interval for men aged 0 in the NICE-2005 group in the Pre cohort.

δ_1 is the difference in mean log-transformed diagnostic interval in the Pre cohort between the NICE-2015-only and NICE-2005 groups.

δ_2 is the Pre-to-Post change in mean log-transformed diagnostic interval attributable to NICE 2015 (see Eq. (A.1)).

Age at diagnosis was entered as a continuous variable.

Table C.3: Pre-to-post difference-in-differences regression analysis results by gender. Pre: 1 Aug 2012 to 31 Dec 2014 ($n=1,205$ men, $n=1,038$ women); Post: 1 Aug 2015 to 31 Dec 2017 ($n=594$ men, $n=423$ women)

Estimand	Men ($n=1,799$)			Women ($n=1,461$)		
	Coeff.	95% CI	p	Coeff.	95% CI	p
δ_0	3.687	3.380 to 3.995	<0.0001	3.992	3.659 to 4.326	<0.0001
δ_1	0.093	-0.054 to 0.240	0.213	0.027	-0.119 to 0.173	0.719
δ_2	0.071	-0.052 to 0.193	0.257	0.115	-0.035 to 0.265	0.132
δ_3	-0.194	-0.447 to 0.059	0.133	-0.085	-0.345 to 0.176	0.523
Age	0.003	-0.002 to 0.007	0.240	0.0003	-0.005 to 0.004	0.881

Table C.4: Results of event-study difference-in-differences analysis, by sex

t^b	Estimate	Men			Women		
		Coeff.	95% CI	P	Coeff.	95% CI	P
	δ_0	3.784	3.584 to 3.985	<0.0001	4.071	3.834 to 4.309	<0.0001
	δ_1	0.034	-0.245 to 0.314	0.810	-0.024	-0.288 to 0.240	0.857
-10	γ_{-10}	0.842	0.633 to 1.050	0.000	0.818	0.574 to 1.063	<0.0001
	λ_{-10}	0.321	-0.040 to 0.682	0.081	0.281	-0.097 to 0.659	0.145
-9	γ_{-9}	-0.039	-0.204 to 0.126	0.642	-0.127	-0.333 to 0.079	0.225
	λ_{-9}	0.247	-0.079 to 0.574	0.137	0.202	-0.136 to 0.540	0.241
-8	γ_{-8}	-0.136	-0.303 to 0.032	0.112	-0.144	-0.336 to 0.048	0.141
	λ_{-8}	0.200	-0.138 to 0.538	0.246	0.280	-0.051 to 0.611	0.097
-7	γ_{-7}	-0.095	-0.251 to 0.061	0.233	-0.259	-0.464 to -0.054	0.013
	λ_{-7}	0.168	-0.184 to 0.520	0.349	0.403	0.072 to 0.735	0.017
-6	γ_{-6}	-0.026	-0.188 to 0.137	0.756	-0.256	-0.469 to -0.043	0.019
	λ_{-6}	-0.066	-0.406 to 0.274	0.702	0.320	-0.024 to 0.664	0.068
-5	γ_{-5}	-0.074	-0.238 to 0.090	0.378	-0.051	-0.255 to 0.153	0.623
	λ_{-5}	0.119	-0.215 to 0.453	0.484	0.173	-0.177 to 0.523	0.332
-4	γ_{-4}	-0.071	-0.237 to 0.095	0.401	-0.056	-0.241 to 0.129	0.553
	λ_{-4}	0.227	-0.101 to 0.555	0.174	-0.037	-0.369 to 0.296	0.828
-3	γ_{-3}	0.005	-0.163 to 0.174	0.951	-0.048	-0.249 to 0.153	0.636
	λ_{-3}	-0.040	-0.396 to 0.315	0.823	0.043	-0.309 to 0.396	0.809
-2	γ_{-2}	-0.153	-0.321 to 0.016	0.075	-0.089	-0.292 to 0.114	0.388
	λ_{-2}	0.091	-0.268 to 0.450	0.618	0.057	-0.307 to 0.421	0.760
-1	γ_{-1}	0.015	-0.162 to 0.192	0.867	-0.019	-0.226 to 0.187	0.853
	λ_{-1}	0.112	-0.234 to 0.459	0.524	0.121	-0.219 to 0.462	0.484
+1	γ_{+1}	0.079	-0.118 to 0.276	0.432	0.114	-0.124 to 0.353	0.346
	λ_{+1}	-0.190	-0.560 to 0.181	0.315	0.036	-0.418 to 0.491	0.875
+2	γ_{+2}	-0.027	-0.221 to 0.167	0.787	0.031	-0.205 to 0.267	0.796
	λ_{+2}	-0.337	-0.779 to 0.105	0.135	-0.299	-0.712 to 0.114	0.155
	Age	0.002	-0.001 to 0.004	0.157	<0.0001	-0.003 to 0.002	0.716

Post-estimation link tests did not reveal any problems with the data specification of the difference-in-differences model.
Time, years relative to 2015.

δ_0 is mean log-transformed diagnostic interval at age 0 in the NICE-2005 group in 2015.

δ_1 is the difference in mean log-transformed diagnostic interval between the NICE-2015-only and NICE-2005 groups in 2015.

γ_t is the difference in mean log-transformed diagnostic interval in NICE-2005 groups between year t and 2015.

λ_t is the difference in mean log-transformed diagnostic interval in year t compared with 2015 that is attributable to NICE 2015.

Table C.5: Pre-to-post difference-in-differences regression analysis results. Pre: 1 Aug 2012 to 31 Dec 2014 ($n=2,243$); Post: 1 Aug 2015 to 31 Dec 2017 ($n=1,017$). Including interaction terms between Pre/Post and age and between Pre/Post and sex

Estimand	Coeff.	95% CI	p
δ_0	3.8	3.5 to 4.1	<0.0001
δ_1	0.06	0.04 to 0.2	0.258
δ_2	0.06	0.4 to 0.5	0.810
δ_3	0.1	0.3 to 0.04	0.125
δ_4	0.08	0.02 to 0.2	0.128
δ_5	0.09	0.07 to 0.3	0.265
δ_6	0.001	-0.003 to 0.005	0.555
δ_7	0.00006	0.007 to 0.006	0.986

Model specification:

$$y = \delta_0 + \delta_1 NICE_{2015} + \delta_2 d_T + \delta_3 d_T \cdot NICE_{2015} + \delta_4 SEX + \delta_5 d_T \cdot SEX + \delta_6 AGE + \delta_7 d_T \cdot AGE + u \quad (C.1)$$

δ_0 is mean log-transformed diagnostic interval for men at age 0 in the NICE-2005 group in the Pre cohort.

δ_1 is the difference in mean log-transformed diagnostic interval in the Pre cohort between the NICE-2015-only and NICE-2005 groups, for men aged 0.

δ_2 is the Pre-to-Post difference in mean log-transformed diagnostic interval in the NICE-2005 group, for men aged 0.

δ_3 is the Pre-to-Post change in mean log-transformed diagnostic interval attributable to NICE 2015.

δ_4 is the female vs male difference in mean log-transformed diagnostic interval in the NICE-2005 group at age 0.

δ_5 is the Pre-to-Post difference in mean log-transformed diagnostic interval attributable to being female.

δ_6 is the mean change in log-transformed diagnostic interval for a unit change in age in men in the Pre group.

δ_7 is the Pre-to-Post difference in mean log-transformed diagnostic interval attributable to a unit change in age.

Table C.6: Event-study difference-in-differences analyses, including interaction terms between time and age, and time and sex

Estimand	Coeff.	95% CI	P
δ_0	3.746	3.292 to 4.200	<0.0001
δ_1	0.006	-0.194 to 0.205	0.956
γ_{-10}	1.222	0.596 to 1.847	<0.0001
λ_{-10}	0.294	0.024 to 0.564	0.033
ζ_{-10}	-0.042	-0.287 to 0.204	0.740
η_{-10}	-0.005	-0.014 to 0.003	0.237
γ_{-9}	0.285	-0.358 to 0.928	0.384
λ_{-9}	0.222	-0.027 to 0.470	0.080
ζ_{-9}	-0.103	-0.328 to 0.122	0.369
η_{-9}	-0.005	-0.013 to 0.004	0.325
γ_{-8}	-0.292	-0.947 to 0.363	0.382
λ_{-8}	0.242	-0.007 to 0.490	0.057
ζ_{-8}	0.024	-0.204 to 0.252	0.835
η_{-8}	0.002	-0.007 to 0.011	0.674
γ_{-7}	0.205	-0.393 to 0.803	0.502
λ_{-7}	0.282	0.043 to 0.521	0.021
ζ_{-7}	-0.061	-0.287 to 0.165	0.598
η_{-7}	-0.005	-0.013 to 0.004	0.267
γ_{-6}	0.019	-0.596 to 0.633	0.952
λ_{-6}	0.116	-0.136 to 0.368	0.365
ζ_{-6}	-0.055	-0.277 to 0.167	0.628
η_{-6}	-0.002	-0.010 to 0.007	0.703
γ_{-5}	0.141	-0.481 to 0.762	0.656
λ_{-5}	0.144	-0.102 to 0.391	0.251
ζ_{-5}	0.049	-0.174 to 0.272	0.668
η_{-5}	-0.003	-0.012 to 0.006	0.485
γ_{-4}	0.151	-0.468 to 0.770	0.632
λ_{-4}	0.103	-0.133 to 0.339	0.392
ζ_{-4}	-0.098	-0.319 to 0.124	0.388
η_{-4}	-0.003	-0.011 to 0.006	0.571
γ_{-3}	0.098	-0.526 to 0.723	0.757
λ_{-3}	0.001	-0.249 to 0.251	0.993
ζ_{-3}	-0.017	-0.258 to 0.224	0.889
η_{-3}	-0.002	-0.011 to 0.007	0.736

γ_{-2}	-0.151	-0.737 to 0.436	0.614
λ_{-2}	0.077	-0.193 to 0.347	0.574
ζ_{-2}	0.048	-0.174 to 0.270	0.672
η_{-2}	0.000	-0.008 to 0.008	0.987
γ_{-1}	0.001	-0.662 to 0.663	0.999
λ_{-1}	0.118	-0.139 to 0.374	0.367
ζ_{-1}	-0.032	-0.260 to 0.196	0.782
η_{-1}	0.000	-0.009 to 0.009	0.968
γ_{+1}	0.094	-0.715 to 0.902	0.820
λ_{+1}	-0.096	-0.397 to 0.206	0.533
ζ_{+1}	0.139	-0.129 to 0.407	0.308
η_{+1}	-0.001	-0.012 to 0.011	0.900
γ_{+2}	0.201	-0.535 to 0.937	0.592
λ_{+2}	-0.316	-0.631 to -0.001	0.049
ζ_{+2}	0.079	-0.213 to 0.371	0.597
η_{+2}	-0.003	-0.014 to 0.007	0.537
v_1	0.002	-0.004 to 0.009	0.467
ψ_1	0.109	-0.070 to 0.287	0.232

Notes for Appendix Table C.6:

δ_0 is mean log-transformed diagnostic interval for men at age 0 in the NICE-2005 group in 2015.

δ_1 is the difference in mean log-transformed diagnostic interval between the NICE-2015-only and NICE-2005 groups for men aged 0 in 2015.

γ_t is the difference in mean log-transformed diagnostic interval in NICE-2005 groups between year t and 2015.

λ_t is the difference in mean log-transformed diagnostic interval in year t compared with 2015 that is attributable to NICE 2015.

ϕ_1 the difference in mean log-transformed diagnostic interval between women and men, at age 0 in NICE-2005 group in 2015.

ζ_t is the additional difference in mean log-transformed diagnostic interval in year t compared with 2015 that is attributable to being female.

v_1 is the change in mean log-transformed diagnostic interval for a unit change in age in men in NICE-2005 group in 2015.

η_t is the additional difference in mean log-transformed diagnostic interval in year t compared with 2015 that is attributable to a unit change in age.

Model specification:

$$\begin{aligned}
y = & \delta_0 + \delta_1 NICE_{2015} + v_1 AGE + \psi_1 SEX \\
& + \sum_{t=-T_L, t \neq 0}^{T_U} \gamma_t d_t + \sum_{t=-T_L, t \neq 0}^{T_U} \lambda_t d_t NICE_{2015} \\
& + \sum_{t=-T_L, t \neq 0}^{T_U} \phi_t d_t SEX \\
& + \sum_{t=-T_L, t \neq 0}^{T_U} \zeta_t d_t AGE + u
\end{aligned}$$

Table C.7: Robustness checks for pre-to-post difference-in-differences analyses: varying the start time of the Pre and Post periods

Estimand	Coeff.	95% CI	P
Pre: 23/7/12–22/12/14 ($n=2,257$); Post: 23/7/15–22/12/17 ($n=1,035$)			
δ_0	3.797	3.577 to 4.017	<0.0001
δ_1	0.062	–0.042 to 0.166	0.241
δ_2	0.094	–0.036 to 0.191	0.059
δ_3	–0.135	–0.322 to 0.051	0.155
Age	0.0009	–0.002 to 0.004	0.549
Sex (f vs m)	0.109	0.029 to 0.189	0.007
Pre: 13/7/12–12/12/14 ($n=2,257$); Post: 13/7/15–12/12/17 ($n=1,055$)			
δ_0	3.778	3.556 to 4.000	<0.0001
δ_1	0.060	–0.045 to 0.164	0.263
δ_2	0.092	–0.005 to 0.189	0.064
δ_3	–0.119	–0.307 to 0.069	0.213
Age	0.001	–0.002 to 0.004	0.450
Sex (f vs m)	0.107	0.028 to 0.186	0.008
Pre: 3/7/12–2/12/14 ($n=2,265$); Post: 3/7/15–2/12/17 ($n=1,062$)			
δ_0	3.774	3.554 to 3.995	<0.0001
δ_1	0.056	–0.048 to 0.161	0.287
δ_2	0.081	–0.016 to 0.178	0.103
δ_3	–0.118	–0.304 to 0.069	0.216
Age	0.001	–0.002 to 0.004	0.409
Sex (f vs m)	0.109	0.029 to 0.188	0.007
Pre: 23/6/12–22/11/14 ($n=2,264$); Post: 23/6/15–22/11/17 ($n=1,073$)			
δ_0	3.790	3.567 to 4.011	<0.0001
δ_1	0.061	–0.043 to 0.164	0.249
δ_2	0.097	0.001 to 0.192)	0.047
δ_3	–0.114	–0.298 to 0.070	0.226
Age	0.001	–0.002 to 0.004	0.533
Sex (f vs m)	0.107	0.0280 to 0.186	0.008

δ_0 is mean log-transformed diagnostic interval for men at age 0 in the NICE-2005 group in the Pre cohort.

δ_1 is the difference in mean log-transformed diagnostic interval in the Pre cohort between the NICE-2015-only and NICE-2005 groups.

δ_2 is the Pre to Post difference in mean log-transformed diagnostic interval in the NICE-2005 group.

δ_3 is the Pre-to-Post change in mean log-transformed diagnostic interval attributable to NICE 2015.

Age was entered as a continuous variable.

Table C.8: Robustness checks for pre-to-post difference-in-differences analyses: varying the start time and width of the Pre and Post periods

Estimand	Coeff.	95% CI	P
Pre: 1/7/13–31/7/13 ($n=86$); Post: 1/7/17–31/7/17 ($n=22$)			
δ_0	3.210	1.979 to 4.440	<0.0001
δ_1	−0.157	−0.729 to 0.416	0.588
δ_2	0.009	−0.464 to 0.483	0.968
δ_3	−0.644	−1.477 to 0.188	0.128
Age	0.009	−0.008 to 0.026	0.302
Sex (f vs m)	0.177	−0.311 to 0.664	0.474
Pre: 1/1/13–31/12/13 ($n=964$); Post: 1/7/17–31/12/17 ($n=294$)			
δ_0	3.649	3.315 to 3.983	<0.0001
δ_1	0.082	−0.076 to 0.240	0.306
δ_2	0.125	−0.018 to 0.267	0.086
δ_3	−0.393	−0.693 to −0.093	0.010
Age	0.002	−0.003 to 0.006	0.472
Sex (f vs m)	0.165	0.044 to 0.286	0.008
Pre: 16/3/14–31/12/14 ($n=647$); Post: 16/3/17–31/12/17 ($n=221$)			
δ_0	3.871	3.474 to 4.267	<0.0001
δ_1	0.163	−0.026 to 0.353	0.091
δ_2	−0.146	−0.295 to 0.024	0.054
δ_3	−0.476	−0.807 to −0.146	0.005
Age	0.0004	−0.005 to 0.006	0.868
Sex (f vs m)	0.169	0.021 to 0.317	0.026
Pre: 1/1/06–22/6/15 ($n=10,767$); Post: 23/6/15–31/12/17 ($n=1,075$)			
δ_0	3.820	3.698 to 3.942	<0.0001
δ_1	0.153	0.104 to 0.203	<0.0001
δ_2	0.098	0.014 to 0.183	0.023
δ_3	−0.215	−0.374 to −0.055	0.008
Age	0.0006	−0.001 to 0.002	0.501
Sex (f vs m)	0.091	0.046 to 0.136	<0.0001

δ_0 is mean log-transformed diagnostic interval for men at age 0 in the NICE-2005 group in the Pre cohort.

δ_1 is the difference in mean log-transformed diagnostic interval in the Pre cohort between the NICE-2015-only and NICE-2005 groups.

δ_2 is the Pre to Post difference in mean log-transformed diagnostic interval in the NICE-2005 group.

δ_3 is the Pre-to-Post change in mean log-transformed diagnostic interval attributable to NICE 2015.

Age was entered as a continuous variable.

Table C.9: Pre-to-post difference-in-differences regression analyses, adjusted for region. Pre: 1 Aug 2012 to 31 Dec 2014; post: 1 Aug 2015 to 31 Dec 2017

Estimand	Coeff.	95% CI	p
δ_0	3.976	3.746 to 4.205	<0.0001
δ_1	0.061	-0.043 to 0.164	0.253
δ_2	0.051	-0.047 to 0.149	0.305
δ_3	-0.144	-0.328 to 0.040	0.126
Age	0.001	-0.002 to 0.004	0.581
Sex (f vs m)	0.110	0.030 to 0.190	0.007
Region	-0.218	-0.306 to -0.129	<0.0001

δ_0 is mean log-transformed diagnostic interval for men at age 0 in the NICE-2005 group in the Pre cohort.

δ_1 is the difference in mean log-transformed diagnostic interval in the Pre cohort between the NICE-2015-only and NICE-2005 groups.

δ_2 is the Pre to Post difference in mean log-transformed diagnostic interval in the NICE-2005 group.

δ_3 is the Pre-to-Post change in mean log-transformed diagnostic interval attributable to NICE 2015.

Age was entered as a continuous variable; region as a binary variable (England vs Wales or Northern Ireland).

Table C.10: Event-study difference-in-differences analysis, adjusting for region (England vs Wales or NI)

Time	Estimand	Coeff.	95% CI	P
	δ_0	4.114	3.956 to 4.272	<0.0001
	δ_1	0.018	-0.179 to 0.215	0.857
-10	γ_{-10}	0.868	0.712 to 1.023	<0.0001
	λ_{-10}	0.295	0.026 to 0.564	0.032
-9	γ_{-9}	-0.027	-0.164 to 0.110	0.698
	λ_{-9}	0.193	-0.052 to 0.438	0.122
-8	γ_{-8}	-0.093	-0.221 to 0.036	0.156
	λ_{-8}	0.221	-0.024 to 0.465	0.077
-7	γ_{-7}	-0.120	-0.244 to 0.005	0.060
	λ_{-7}	0.256	0.020 to 0.491	0.034
-6	γ_{-6}	-0.087	-0.220 to 0.046	0.199
	λ_{-6}	0.103	-0.144 to 0.349	0.414
-5	γ_{-5}	-0.024	-0.154 to 0.106	0.718
	λ_{-5}	0.140	-0.103 to 0.383	0.260
-4	γ_{-4}	-0.029	-0.151 to 0.093	0.638
	λ_{-4}	0.083	-0.150 to 0.316	0.485
-3	γ_{-3}	0.010	-0.116 to 0.136	0.879
	λ_{-3}	0.000	-0.246 to 0.246	0.999
-2	γ_{-2}	-0.086	-0.221 to 0.049	0.211
	λ_{-2}	0.064	-0.206 to 0.334	0.642
-1	γ_{-1}	0.019	-0.115 to 0.152	0.783
	λ_{-1}	0.107	-0.146 to 0.359	0.407
+1	γ_{+1}	0.078	-0.072 to 0.229	0.308
	λ_{+1}	-0.124	-0.427 to 0.180	0.425
+2	γ_{+2}	-0.056	-0.201 to 0.089	0.450
	λ_{+2}	-0.318	-0.628 to -0.008	0.044
	Age	0.000	-0.001 to 0.002	0.701
	Sex (f vs m)	0.094	0.050 to 0.138	<0.0001
	Region	-0.309	-0.359 to -0.259	<0.0001

δ_0 is mean log-transformed diagnostic interval for men aged 0 in the NICE-2005 group in 2015, in England.

δ_1 is the difference in mean log-transformed diagnostic interval between the NICE-2015-only and NICE-2005 groups in 2015, for men aged 0 in England.

γ_t is the difference in mean log-transformed diagnostic interval in NICE-2005 groups between year t and 2015 in England for men aged 0.

λ_t is the difference in mean log-transformed diagnostic interval in year t compared with 2015 that is attributable to NICE 2015.

Age at diagnosis was entered as a continuous variable; Region was entered as a binary variable (England vs Wales or NI).

Table C.11: Pre-to-post difference-in-differences regression analyses on England-only data, adjusted for deprivation. Pre: 1 Aug 2012 to 31 Dec 2014; post: 1 Aug 2015 to 31 Dec 2017

Estimand	Coeff.	95% CI	p
δ_0	3.6	3.3 to 3.9	<0.0001
δ_1	0.08	−0.06 to 0.2	0.283
δ_2	0.07	−0.08 to 0.2	0.376
δ_3	−0.03	−0.3 to 0.2	0.810
Age	0.001	−0.003 to 0.005	0.491
Sex (f vs m)	0.08	−0.02 to 0.2	0.113
Deprivation	0.1	−0.03 to 0.2	0.147

δ_0 is mean log-transformed diagnostic interval for men at age 0 in the NICE-2005 group in the Pre cohort.

δ_1 is the difference in mean log-transformed diagnostic interval in the Pre cohort between the NICE-2015-only and NICE-2005 groups.

δ_2 is the Pre to Post difference in mean log-transformed diagnostic interval in the NICE-2005 group. δ_3 is the Pre-to-Post change in mean log-transformed diagnostic interval attributable to NICE 2015.

Age at diagnosis was entered as a continuous variable; deprivation as a binary variable (most deprived quintiles 4 and 5 vs least deprived quintiles 1-3).

Table C.12: Event-study difference-in-differences analysis of England-only data, adjusting for deprivation (most vs least)

Time	Estimand	Coeff.	95%CI	P
	δ_0	3.691	3.484 to 3.899	<0.0001
	δ_1	-0.084	-0.373 to 0.205	0.568
-10	γ_{-10}	0.959	0.738 to 1.181	<0.0001
	λ_{-10}	0.428	0.045 to 0.810	0.028
-9	γ_{-9}	0.005	-0.165 to 0.175	0.955
	λ_{-9}	0.317	-0.023 to 0.658	0.068
-8	γ_{-8}	-0.068	-0.234 to 0.099	0.425
	λ_{-8}	0.305	-0.039 to 0.650	0.082
-7	γ_{-7}	-0.107	-0.269 to 0.056	0.198
	λ_{-7}	0.399	0.068 to 0.731	0.018
-6	γ_{-6}	-0.052	-0.232 to 0.129	0.574
	λ_{-6}	0.198	-0.164 to 0.560	0.283
-5	γ_{-5}	-0.008	-0.179 to 0.163	0.926
	λ_{-5}	0.227	-0.102 to 0.556	0.176
-4	γ_{-4}	-0.058	-0.219 to 0.104	0.483
	λ_{-4}	0.198	-0.137 to 0.534	0.245
-3	γ_{-3}	0.013	-0.150 to 0.177	0.873
	λ_{-3}	0.076	-0.272 to 0.424	0.668
-2	γ_{-2}	-0.116	-0.297 to 0.064	0.207
	λ_{-2}	0.204	-0.185 to 0.593	0.304
-1	γ_{-1}	0.114	-0.064 to 0.292	0.209
	λ_{-1}	0.189	-0.168 to 0.545	0.298
+1	γ_{+1}	0.195	-0.010 to 0.400	0.062
	λ_{+1}	0.102	-0.325 to 0.530	0.638
+2	γ_{+2}	0.075	-0.135 to 0.286	0.483
	λ_{+2}	-0.226	-0.720 to 0.268	0.370
	Age	0.001	-0.001 to 0.003	0.437
	Sex (f vs m)	0.085	0.028 to 0.142	0.003
	Deprivation	0.069	0.002 to 0.136	0.043

δ_0 is mean log-transformed diagnostic interval for men aged 0 in the NICE-2005 group in 2015, in the least deprived quintiles (4 & 5).

δ_1 is the difference in mean log-transformed diagnostic interval between the NICE-2015-only and NICE-2005 groups in 2015, for men aged 0 in quintiles 4 & 5.

γ_t is the difference in mean log-transformed diagnostic interval in NICE-2005 groups between year t and 2015 for men aged 0 in quintiles 4 & 5.

λ_t is the difference in mean log-transformed diagnostic interval in year t compared with 2015 that is attributable to NICE 2015.

Age at diagnosis was entered as a continuous variable.